

Begriffsrelationen im Kategoriensystem der Wikipedia

Entwicklung eines Relationeninventars zur kollaborativen Anwendung

Bachelorarbeit

Studiengang Bibliothekswesen

Fakultät für Informations- und Kommunikationswissenschaften

Fachhochschule Köln

vorgelegt von:

Alexander Meyer

Venloer Str. 500

50825 Köln

Matrikel-Nr.: 11 06 32 26

am 30. August 2010 bei Prof. Dipl.-Math. Winfried Gödert

Zusammenfassung

Neben den bekannten Hyperlinks in Artikeltexten verfügt die Online-Enzyklopädie Wikipedia mit ihrem Kategoriensystem über ein weiteres Mittel zur Herstellung von Relationen zwischen Artikeln. Jeder Artikel ist einer oder mehreren Kategorien zugeordnet, die ihrerseits anderen Kategorien zugeordnet sind. Auf diese Weise entsteht eine systematische Ordnung von Artikeln und Kategorien.

Betrachtet man nur die Artikel- und Kategoriennamen sowie diese Relationen, so stellt das Kategoriensystem ein gemeinschaftlich erstelltes Begriffssystem dar, das sämtliche von der Wikipedia abgedeckten Themenbereiche umfasst, jedoch – technisch betrachtet – ausschließlich hierarchische Relationen enthält. Aufgrund des Fehlens eines differenzierten Regelwerks zur Kategorisierung ist das Kategoriensystem derzeit jedoch inkonsistent, daneben sind, bedingt durch das Vorhandensein lediglich eines Relationstyps, viele Relationen wenig aussagekräftig. Dennoch besteht das Potenzial zur Schaffung eines stark und differenziert relationierten Begriffssystems aus dem bestehenden Kategoriensystem heraus.

Die vorliegende Arbeit diskutiert die Anwendungsmöglichkeiten eines solchen Begriffssystems und die Option seiner gemeinschaftlichen Entwicklung aus dem bestehenden Vokabular des Kategoriensystems, mithin also der gemeinschaftlichen Relationierung von Begriffen anhand eines differenzierten Relationeninventars. Ausgehend von den Kategorien „Theater“ und „Jagd“ der deutschsprachigen Wikipedia wird ein hierarchisch strukturiertes Relationeninventar definiert, das sowohl spezifische als auch allgemeine Relationstypen enthält und damit die Möglichkeit zur Übertragung auf andere Gegenstandsbereiche bietet. Sämtliche Artikel- und Kategoriennamen, die unterhalb jener Kategorien erscheinen, werden unter Verwendung der neu entwickelten Relationstypen als Deskriptoren relationiert.

Schlagwörter: Wikipedia, semantische Relation, Thesaurus, Ontologie, Relationstyp

Inhaltsverzeichnis

1	Einleitung	6
2	Relationstypen in Dokumentationssprachen	10
2.1	Klassische Relationstypen	11
2.2	Erweiterte Relationeninventare	13
3	Das Kategoriensystem der Online-Enzyklopädie Wikipedia	16
3.1	Das Wiki-Prinzip und die Geschichte der Wikipedia	16
3.2	Entstehung und Entwicklungsstand des Kategoriensystems	18
3.3	Typologie des Kategoriensystems als Begriffssystem und derzeitige Problemstellungen	23
3.4	Anwendungen des Kategoriensystems außerhalb der Wikipedia	28
3.5	Fazit und Ausblick	31
4	Grundüberlegungen zur Entwicklung eines differenzierten Relationen- inventars	32
4.1	Formulierung einer Struktur anhand eines Beispiels	32
4.2	Nutzen eines differenzierten Relationeninventars	38
4.3	Entwicklung eines Relationeninventars anhand bestehenden Vokabulars	39
5	Entwicklung eines Relationeninventars zur gemeinschaftlichen Anwen- dung im Kategoriensystem der Wikipedia	41
5.1	Wikipedia-spezifische Grundüberlegungen	41
5.2	Beschreibung sämtlicher Arbeitsschritte	44
5.3	Exemplarische Relationierung des Gegenstandsbereichs Theater der deutschsprachigen Wikipedia	45
5.3.1	Derzeitige Struktur	45
5.3.2	Intellektuelles Vorgehen	47
5.3.3	Technisches Vorgehen	49
5.3.4	Das neue Relationeninventar	52
5.3.5	Exemplarische Darstellung der neuen Strukturen mit Graphviz	56
5.4	Anwendung des Inventars auf den Gegenstandsbereich Jagd	61
5.5	Entwicklung eines hierarchisch strukturierten themenübergreifenden Relationeninventars	63
6	Schluss	67
	Literatur	69

Anhang	73
A Relationeninventar für den Gegenstandsbereich Theater	73
B Relationeninventar für den Gegenstandsbereich Jagd	79
C Visualisierungen mit Graphviz	83
C.1 Quelltexte zu den Abbildungen 6 und 7	83
C.2 Weitere Beispiele	86
D Skripte für „bash“	91
E Verzeichnisstruktur der beigelegten CD-ROM	103

Abbildungsverzeichnis

1	Der Kategorie „Jagd“ untergeordnete Kategorien – nicht Artikel – im Kategoriensystem der deutschsprachigen Wikipedia (Darstellung mit Catgraph als Wolke)	21
2	Der Kategorie „Jagd“ übergeordnete Kategorien im Kategoriensystem der deutschsprachigen Wikipedia (Darstellung mit Catgraph als Baum)	22
3	Die Position des Artikels „Agnus Dei“ im Kategoriensystem der deutschsprachigen Wikipedia	29
4	Direkt untergeordnete Kategorien der Kategorie „Theater“ im bisherigen Stand	46
5	Direkt untergeordnete Kategorien und Artikel der Kategorie „Theaterbegriff“ im bisherigen Stand	48
6	Das semantische Umfeld des Begriffs Theaterregie in den neuen Strukturen	59
7	Die Unterbegriffe des Begriffs Theater, deren Unterordnung auf dem Aspekt des behandelten Stoffs <i>und</i> einem historischen Bezug gründet, in den neuen Strukturen	60
8	Das semantische Umfeld der Begriffe Theaterregie und Operninszenierung (manuell anhand der Skriptausgaben zusammengefügt)	62
9	Die Unterbegriffe des Begriffs Theater, deren Unterordnung auf einem geographischen <i>und</i> einem zeitlichen Bezug gründet, in den neuen Strukturen	86
10	Das semantische Umfeld des Begriffs Drama in den neuen Strukturen (nur neun Relationstypen)	87
11	Das semantische Umfeld des Begriffs Theaterfigur in den neuen Strukturen	88
12	Das semantische Umfeld des Begriffs Jagd in den neuen Strukturen (nur neun Relationstypen)	89
13	Das semantische Umfeld des Begriffs Jagdwaffe in den neuen Strukturen	90

1 Einleitung

Die Wikipedia ist eine 2001 gegründete, im Internet frei zugängliche Enzyklopädie. Ihre Texte werden von ehrenamtlichen und zum Teil anonym oder pseudonym agierenden Internetnutzern geschrieben und können von jedem Internetnutzer geändert und erweitert werden.¹ Es existieren Fassungen in mehr als 260 Sprachen beziehungsweise Dialekten, wobei 32 Sprachfassungen mehr als 100.000 Artikel (Lemmata mit den dazugehörigen Texten) enthalten.² Die Artikeltexte sind als Hypertext organisiert: Ein Wort oder eine Phrase aus dem Text, die in ihrer Grundform ebenfalls als Lemma vorhanden sind, können mit diesem verknüpft werden. Es wird dadurch ein Hyperlink zum jeweiligen Artikel angelegt. Auf diese Weise entsteht ein Netz aus miteinander verwobenen Artikeln, wobei sämtliche Verknüpfungen nur in eine Richtung gelten – nämlich aus dem Text eines Artikels heraus zu einem anderen Artikel – und ihre Natur formal nicht näher spezifiziert ist. Daneben ist das Auffinden von Lemmata auch durch ihr Eingeben in eine Suchmaske möglich.

Gemessen an der Zahl der Artikel, hat die englischsprachige Fassung mit etwa 3.000.000 im Juni 2010 den größten Umfang, gefolgt von der deutschsprachigen mit etwa 1.000.000 zum gleichen Zeitpunkt. Eine solche Größe legt den Wunsch nahe, die Lemmata nicht nur durch die genannten undifferenzierten und nicht systematischen Verknüpfungen aus Texten heraus beziehungsweise durch das Eingeben von Suchbegriffen auffindbar zu machen, sondern auch eine systematische Sortierung von Lemmata und damit Artikeln zu entwickeln. Bereits seit 2004 existiert hierfür das Konzept der Kategorien.³ Kategorien heißen in diesem Zusammenhang natürlichsprachliche Begriffsbezeichnungen, die mit Schlagwörtern aus Dokumentationssprachen vergleichbar sind. Jeder Artikel wird einer oder mehrerer sogenannter Kategorien zugeordnet, dieser Vorgang kann mit der Vergabe von Schlagwörtern verglichen werden. Kategorien werden ihrerseits anderen Kategorien zugeordnet. So entsteht – technisch betrachtet – eine hierarchische Ordnung von Kategorien und Artikeln, an deren Spitze derzeit eine einzige Hauptkategorie und am Ende der Hierarchien die Artikel stehen.

¹Vgl. [Wikipedia 2010a], den Eintrag zur Wikipedia selbst in der deutschsprachigen Wikipedia. Zur Geschichte der Wikipedia siehe Abschnitt 3.1 ab S. 16 dieser Arbeit. Der Begriff der Enzyklopädie ist zur Beschreibung der Wikipedia zwar problematisch, da sie vielmehr den Charakter eines Lexikons beziehungsweise eines enzyklopädischen Lexikons hat, wird aber in Übereinstimmung mit der Selbstdarstellung der Wikipedia in dieser Arbeit übernommen.

²Die Umfangsangaben entstammen [Wikipedia 2010b]. Bei den einzelnen Sprachfassungen handelt es sich nicht um Übersetzungen einer zentralen Enzyklopädie, sondern sie haben vielmehr eigene Inhalte, werden jedoch mit dem gleichen Anspruch der Schaffung einer frei zugänglichen gemeinschaftlich entwickelten Enzyklopädie geführt. Wenn in dieser Arbeit von „der Wikipedia“ im Allgemeinen die Rede ist, so sind sämtliche Sprachfassungen gemeint. Wird dagegen eine Sprachfassung genannt, etwa „die deutschsprachige Wikipedia“, so bezieht sich die Aussage nur auf jene Fassung.

³Vgl., auch zur Problematik der Verwendung des Begriffs der Kategorie in diesem Zusammenhang, Abschnitt 3 ab S. 16 dieser Arbeit.

Aus dokumentarischer Sicht zeigen sich im Kategoriensystem der Wikipedia also Parallelen zu Begriffssystemen beziehungsweise Dokumentationssprachen, die Relationen zwischen natürlichsprachlichen Bezeichnungen ausweisen, etwa Taxonomien, Thesauri und Ontologien. Die vorliegende Arbeit begreift die Namen von Kategorien und Artikeln, sofern sie geistige Einheiten oder Gegenstände bezeichnen⁴, als Begriffe eines solchen Begriffssystems – dies schließt, wenn nötig, ihre sprachliche Vereinheitlichung ein – und die Zuordnung eines Artikels oder einer Kategorie in eine andere als die Ausweisung einer, wie auch immer gearteten, Relation zwischen ihnen: Wird beispielsweise die Kategorie „Drama“ in die Kategorie „Theater“ einsortiert, so bedeutet dies, dass zwischen den Begriffen Drama und Theater eine zu diesem Zeitpunkt noch nicht näher bestimmte Beziehung besteht.

Diese Arbeit untersucht exemplarisch anhand zweier Gegenstandsbereiche die so bestehenden Relationen auf der semantischen Ebene. Dabei werden die vorhandenen Relationen, die ja schlicht durch die Einordnung eines Artikels oder einer Kategorie in eine andere Kategorie entstanden sind, in ihrer semantischen Natur untersucht und differenziert typisiert, und es wird ein differenziertes, über die allgemeine hierarchische und die allgemeine assoziative Relation hinausgehendes Inventar von Relationstypen entwickelt, das für weitere Relationierungen und Relationentypisierungen im Kategoriensystem eingesetzt werden kann.

Die Idee der Entwicklung eines differenziert relationierten Begriffssystems aus dem Kategoriensystem ergibt sich vor allen Dingen aus dessen Größe und dem damit einhergehenden Wunsch, Möglichkeiten der Navigation und der Suche einzuführen, die über eine alphabetische Anzeige aller einer Kategorie zugeordneten Kategorien und Artikel hinausgehen. So kennt die deutschsprachige Wikipedia nach dem Stand vom 3. Juni 2010 147.578 Kategorien sowie 4.505.156 Verknüpfungen, also Zuordnungen, zwischen ihnen beziehungsweise zwischen ihnen und Artikeln.⁵

Da die vorhandenen Begriffsrelationen, die also schlicht beim Aufbau des Kategoriensystems entstehen, kollaborativ geschaffen wurden, liegt es nahe, auch die Relationierungsarbeit anhand differenzierter Relationstypen gemeinschaftlich durchzuführen. Dabei wäre wahrscheinlich, da die Zahl der Relationen bereits sehr hoch ist, keine komplette Neurelationierung aller Kategorien und Artikel nötig, sondern es kann, wo es möglich ist, direkt an den bestehenden Relationen gearbeitet werden. Das Einfügen oder Löschen von Relationen wäre nur dort notwendig, wo eine sinnvolle Relation fehlt oder eine bestehende Relation in keinerlei Hinsicht sinnvoll erscheint. Vielmehr benötigt ein sinnvoll anwendbares Begriffssystem nicht eine höhere oder niedrigere Zahl an Relationen, sondern differenziertere Relationen. Technisch

⁴Dies schließt diejenigen Artikel ein, die einen Begriff beschreiben – quantitativ ist dies der größte Teil –, und diejenigen aus, für die dies nicht gilt, beispielsweise den Artikel „Liste von Operetten“.

⁵In diese Zahlen sind auch die sogenannten Wartungskategorien einbezogen, die nicht der inhaltlichen Erschließung dienen, sondern der Markierung verbesserungsbedürftiger Artikel. Diese sind jedoch sehr wenige.

betrachtet werden, wie bereits erwähnt, die existierenden Relationen aufgrund der in dieser Hinsicht begrenzten Möglichkeiten der eingesetzten Software als hierarchische bezeichnet. Dieser Relationstyp ist wenig aussagekräftig, denn es gibt eine Vielzahl von Aspekten, auf die sich die Über- und Unterordnung gründen kann. Im vorliegenden System ist zudem unklar, ob jede hierarchisch genannte Beziehung wirklich eine solche ist und nicht vielmehr im Sinne eines Thesaurus eine assoziative, die mangels anderer technischer Möglichkeiten als hierarchisch angegeben wurde.

Ein Begriffssystem, das diese Problematik beseitigen möchte, benötigt also zunächst ein differenziertes Inventar an Relationstypen, das mehr als einen Typen enthält. Ein solches Inventar muss im Vorfeld festgelegt werden und kann daraufhin zur Relationierung eingesetzt werden.⁶ Da das Inventar nicht ausschließlich von informationswissenschaftlich ausgebildeten Fachleuten, sondern von einer nicht näher umrissenen Gemeinschaft von Internetnutzern zur kollaborativen Relationierung verwendet werden soll, muss es möglichst allgemeinverständlich sein. Ebenso muss es sich für eine Darstellung seiner selbst sowie mit ihm entwickelter Begriffsrelationen innerhalb einer graphischen Oberfläche sowie in Listenform eignen, analog zu den derzeitigen Darstellungsmöglichkeiten des Wikipedia-Kategoriensystems.⁷ Nicht nur das Relationeninventar, sondern auch das neue Begriffssystem sollen für eine Verwendung durch Menschen, nicht durch Maschinen, konzipiert werden.

Ein Vorgehen zur Entwicklung eines hierfür geeigneten Inventars sowie ein so tatsächlich entwickeltes Inventar beschreibt [Boteram 2008]. Boteram geht dabei deduktiv vor und analysiert zunächst die bereits bestehenden Relationen in einem bereits existierenden, aber formal wenig differenziert relationierten Vokabular – dem Teil zum Gegenstandsbereich des Theaters der Schlagwortnormdatei (SWD). Anhand dieser Untersuchungen entwickelt er ein hierarchisch strukturiertes Relationeninventar und relationiert das vorhandene Vokabular unter Verwendung dieses Inventars.

Dieses Vorgehen wird in der vorliegenden Arbeit übernommen: Exemplarisch wird, ausgehend von der Kategorie „Theater“ der deutschsprachigen Wikipedia, die Natur sämtlicher Relationen zwischen darunterliegenden Kategorien und von Kategorien zu Artikeln untersucht. Davon ausgehend wird durch Abstraktion von jeder konkreten Relation ein hierarchisch strukturiertes Relationeninventar, zunächst spezifisch für den Gegenstandsbereich Theater, definiert.⁸ Im Rahmen der Entwicklung des

⁶Um die Konsistenz der Dokumentationssprache zu gewährleisten, darf das Relationeninventar naturgemäß, im Gegensatz zu der Artikeln und Kategorien in der Wikipedia, nicht jederzeit möglichen Änderungen durch jeden Internetnutzer unterliegen, sondern muss vor dem Beginn der Relationentypisierung feststehen. Die Relationen an sich sollen jedoch, wie das derzeit bestehende Kategoriensystem, entsprechend der kollaborativen Natur der Wikipedia jederzeit änderbar sein.

⁷Zu den derzeitigen Darstellungsmöglichkeiten der Relationen vgl. S. 20 dieser Arbeit.

⁸Da das Wikipedia-Kategoriensystem bei gleicher thematischer Abdeckung weit mehr Begriffe und Relationen als die Schlagwortnormdatei enthält, liegt es nahe, dass es einen von Grund auf anderen Aufbau hat und somit die im Kategoriensystem bestehenden Relationen zum Teil anderer Natur als diejenigen in der Schlagwortnormdatei sind. Deswegen kann das von Boteram defi-

Inventars werden zudem sämtliche Beziehungen, die im genannten Gegenstandsbereich erscheinen, typisiert. Darüber hinaus werden, wenn nötig, die Bezeichnungen von Kategorien und Artikeln in diesem Bereich sprachlich vereinheitlicht, sodass sie Deskriptoren darstellen.

Das entwickelte theaterspezifische Inventar wird daraufhin auf einen vollkommen anderen Gegenstandsbereich, ausgehend von der Kategorie „Jagd“, angewandt. Durch diesen Arbeitsschritt wird darauf geschlossen, welche der zunächst theaterspezifisch entwickelten Relationstypen sich für eine allgemeine themenübergreifende Anwendung eignen und welche nur im Gegenstandsbereich Theater eine Rolle spielen. Anhand dieser Erkenntnisse wird am Schluss ein erster Vorschlag eines themenübergreifenden und doch differenzierten und feinen Relationeninventars entwickelt, das, wenn nötig, an allen Stellen themenspezifisch erweitert werden kann.

Innerhalb dieser Arbeit entstehen also zunächst zwei themenspezifische Relationeninventare für zwei Gegenstandsbereiche sowie zwei auf den bestehenden Teilen des Kategoriensystems basierende Vokabulare, die mit den jeweiligen Inventaren relationiert sind. Daraufhin wird ein allgemein anwendbares Relationeninventar mit der Möglichkeit themenspezifischer Erweiterungen entwickelt.

Diese Arbeit steht somit in einer Reihe von Projekten, die sich mit innerhalb der Wikipedia entstandenen und vorhandenen semantischen Relationen zwischen Begriffen oder anderem befassen. Dabei beschränkt sie sich auf die im Kategoriensystem vorhandenen Begriffsrelationen, wobei jedoch, im Gegensatz zu anderen Arbeiten, insbesondere die Allgemeinbegriffe ins Auge gefasst und ihre semantischen Relationen differenziert untersucht werden. Andere Projekt wie DBpedia oder YAGO⁹ fokussieren hingegen konkrete Instanzen und Individualbegriffe und sammeln Informationen über diese, benutzen zur Extraktion der Informationen jedoch zum Teil nicht nur das Kategoriensystem, sondern auch unter anderem die Artikeltexte.

nierte Inventar nicht übernommen werden, sondern muss anhand seiner Methode neu entwickelt werden.

⁹Vgl. Abschnitt 3.4 dieser Arbeit.

2 Relationstypen in Dokumentationssprachen

Vor der Entwicklung eines Relationeninventars steht naturgemäß eine Betrachtung bisher in Begriffssystemen erscheinender Relationstypen. Diese werden im folgenden Abschnitt aus dokumentarischer Sicht anhand verschiedener Formen von Dokumentationssprachen untersucht. Die so gewonnenen Erkenntnisse sind auf Begriffssysteme jeglicher Art übertragbar.

Auf eine ausführliche Darstellung des Zwecks von Dokumentationssprachen und sich daraus ergebende Gestaltungsprinzipien wird an dieser Stelle verzichtet. Hierfür sei auf das Kapitel 7 von [Bertram 2005] (S. 127–147) sowie auf das Kapitel 2 von [Boteram 2008] (S. 8–21) hingewiesen. Dargelegt sei jedoch die folgende Definition des Begriffs der Dokumentationssprache von Boteram, an die sich auch diese Arbeit hält:

„Dokumentationssprachen sind künstliche Sprachen, welche zur Erschließung, Ordnung und zum Retrieval, d. h. dem gezielten Wiederauffinden von Dokumenteninhalten oder deren Repräsentanten innerhalb eines Dokumentationsystems verwendet werden.“ ([Boteram 2008], S. 10.)

Dabei bedeutet der Begriff „künstlich“ in diesem Fall, dass ihre Entwicklung im Gegensatz zur Alltagssprache bewusst und planvoll geschieht, nicht jedoch, dass in ihnen keine natürliche Sprache vorkäme. Insbesondere arbeiten Thesauri und semantische Netze in der Regel mit natürlichsprachlichen Begriffen und stellen die Beziehungen zwischen ihnen dar. Auf dieses Grundprinzip stützt sich auch das Begriffssystem, deren Entwicklung in dieser Arbeit skizziert wird.

Die Benennung der Zwecke einer Dokumentationssprache zur Erschließung und zum Wiederauffinden (Retrieval) von Dokumenten legt nahe, dass sich die Gestaltung von Dokumentationssprachen pragmatisch an diesen Zwecken orientieren soll. Dokumentationssprachen dienen also nicht in erster Linie der Repräsentation von Wissen über einen Gegenstand als Selbstzweck oder um diesen Gegenstand vollständig zu definieren oder um aus dem Wissen Schlussfolgerungen zu ziehen. Diese Möglichkeiten bleiben immer der Eignung zur Dokumenterschließung und zum Retrieval untergeordnet.

Betrachtet man bekannte Typen von Dokumentationssprachen im Hinblick auf die Zahl der in ihnen erscheinenden Relationstypen, so ergibt sich diese Rangfolge:¹⁰

1. Klassifikationen und Taxonomien: ausschließlich hierarchische Relationen.
2. Thesauri: hierarchische, assoziative und Äquivalenzrelationen.

¹⁰Diese Rangfolge ergibt sich aus der Abgrenzung der Begriffe Klassifikation, Thesaurus und Ontologie in [Bertram 2005], S. 261. Die auch hier zugrunde liegende Definition des Begriffs der Taxonomie im Hinblick auf die vorkommenden Relationen erscheint in [Boteram 2008], S. 66.

3. Ontologien (semantische Netze): eine unbegrenzte Zahl an Relationstypen.

Während die Begriffe der Klassifikation und des Thesaurus in der Literatur klar umrissen sind beziehungsweise zu ihnen Normen existieren¹¹, ist eine genaue Definition der Begriffe der Ontologie und des semantischen Netzes sowie der Abgrenzung zwischen ihnen nach wie vor schwierig und hat sich in der Literatur noch nicht durchgesetzt.¹² Deswegen muss auf das dieser Arbeit zugrunde liegende Verständnis beider Begriffe hingewiesen werden:

Da beide Begriffe in dieser Arbeit in erster Linie im Hinblick auf die in ihnen enthaltenen Relationstypen in Abgrenzung zu Klassifikationen und Thesauri betrachtet werden, werden sie in diesem Kontext als synonym angesehen. Zu betonen ist, dass sich ihre Entwicklung nicht aus dem Bereich der Dokumentation, sondern aus den Bereichen der „Kognitionspsychologie und der Neurowissenschaften“ ([Boteram 2008], S. 30.) vollzog und sie der formalen Modellierung und Repräsentation von Wissen über einen Gegenstandsbereich dienen. Dabei wird in der Regel auch die Möglichkeit benannt, Maschinen aus in einer Ontologie angegebenen Begriffsbeziehungen Schlussfolgerungen ziehen zu lassen, die in der Ontologie nicht explizit formuliert sind.¹³ Dies unterscheidet sie von dokumentarischen Klassifikationen und Thesauri, die von vornherein zur Dokumenterschließung konzipiert werden.

Dennoch liegt der Schluss nahe, das zu tun, was Boteram das „[Herausgreifen] einzelne[r] Aspekte semantischer Technologien [...] für die Gestaltung von Dokumentationssprachen und das Information Retrieval“ ([Boteram 2008], S. 21) nennt, also die Übertragung einzelner Merkmale von Ontologien in Dokumentationssprachen. Diese Idee ergibt sich naturgemäß vor allen Dingen aus der Tatsache, dass Ontologien Begriffsbeziehungen jeglicher Art enthalten können. So entstandene Dokumentationssprachen werden in Ermangelung eines spezialisierten Begriffs aus dem dokumentarischen Umfeld – der etwa „komplexer Thesaurus“ lauten könnte – in der Regel auch als Ontologien beziehungsweise semantische Netze bezeichnet.¹⁴ So hält sich auch diese Arbeit an die schlichte Definition einer Ontologie als ein Begriffssystem, das differenziertere Relationstypen als ein Thesaurus enthält.

2.1 Klassische Relationstypen

Betrachtet man die Tatsache, dass Klassifikationen und Thesauri bereits seit Jahrzehnten beziehungsweise noch länger in der Dokumenterschließung eingesetzt wer-

¹¹Vgl. Kapitel 7 bis 12 von [Bertram 2005].

¹²Vgl. [Boteram 2008], S. 29.

¹³Nähere Erläuterungen zum Begriff und zu Anwendungsmöglichkeiten von Ontologien, ebenso weitere Literaturhinweise enthält [Wikipedia 2010c], der Artikel zum Lemma in der deutschsprachigen Wikipedia.

¹⁴So wird beispielweise auch die in [Boteram 2008] entwickelte, differenziert relationierte Dokumentationssprache als Ontologie bezeichnet.

den, so liegt es nahe, die in ihnen enthaltenen Begriffsrelationen als die klassischen Begriffsrelationen in Dokumentationssprachen zu bezeichnen. Diese sind also:

- die hierarchische Relation
- die Äquivalenzrelation
- die assoziative (nicht-hierarchische) Relation

Die hierarchische Relation bedeutet eine wie auch immer geartete Über- und Unterordnung zwischen zwei Begriffen, also die Vorstellung, dass der übergeordnete Begriff den untergeordneten in irgendeiner Hinsicht einschließt.¹⁵ Sie wird in der Thesaurusliteratur meist spezifiziert, wobei in der Regel zwischen den beiden im Folgenden erstgenannten Typen unterschieden¹⁶ und manchmal auch der dritte Typ eingeführt¹⁷ wird:

1. Die generische Relation: „Ein Begriff gilt dann einem anderen als generisch untergeordnet, wenn zu seiner Bestimmung alle Merkmale benutzt werden müssen wie für die Bestimmung des übergeordneten, der untergeordnete sich aber in (mindestens) einem weiteren spezifischen Merkmal vom übergeordneten unterscheidet.“ ([Wersig 1978], S. 127). Beispiele hierfür sind die Beziehungen zwischen dem Oberbegriff Turm und den Unterbegriffen Wachturm und Aussichtsturm. Diese Beziehung ist die in Klassifikationen und biologische Taxonomien in der Regel auftretende.¹⁸
2. Die partitive Relation: Diese Relation liegt vor, wenn der Unterbegriff einen Teil des durch den Oberbegriff Repräsentierten bezeichnet. Beispiele hierfür sind die Beziehungen zwischen dem Oberbegriff Turm und den Unterbegriffen Wendeltreppe und Aussichtsplattform.
3. Die Instanzrelation: Diese Relation liegt vor, wenn der Oberbegriff ein Allgemeinbegriff und der Unterbegriff ein Individualbegriff ist, der eine *individuelle* Ausprägung des durch den Oberbegriff Repräsentierten darstellt. Beispiele hierfür sind die Beziehungen zwischen dem Oberbegriff Turm und den Unterbegriffen Eiffelturm und Schiefer Turm von Pisa.

Während beispielsweise der US-amerikanische Standard zur Gestaltung von einsprachigen kontrollierten Vokabularen – [NISO 2005] – vorschlägt, zwischen diesen drei hierarchischen Relationstypen innerhalb solcher Vokabulare auch formal zu unterscheiden, indem etwa eine generische Unterordnung mit „NTG“, eine partitive mit

¹⁵Vgl. [Wersig 1978], S. 125.

¹⁶Vgl. [Wersig 1978], S. 127–132.

¹⁷Vgl. [Aitchison 2000], S. 54–60.

¹⁸Die Aussagen im letzten Satz sind [Wersig 1978], S. 127, beziehungsweise [Aitchison 2000], S. 56, entnommen.

„NTP“ bezeichnet wird¹⁹, verzichten viele Thesauri auf eine solche Spezifizierung und fassen sämtliche hierarchische Beziehungen unter einem Relationstypen zusammen.

Die Äquivalenzrelation zeigt an, dass zwei Begriffe im Kontext der Dokumentationssprache als äquivalent anzusehen sind. Hierbei wird in der Regel einem den Vorrang gegeben, dieser ist der Vorzugsbegriff, die anderen bekommen den Status von Synonymen. Dabei handelt es sich oft nicht um Synonymie im sprachwissenschaftlichen Sinne, sondern es werden die beiden Begriffe vielmehr pragmatisch für den Zweck der Erschließung und des Retrievals als austauschbar angesehen. Beispiele für verschiedene Typen von Äquivalenzrelationen liefern [Wersig 1978] (S. 112–114) und [Aitchison 2000] (S. 50–54), formal wird jedoch in den meisten Thesauri nicht zwischen diesen unterschieden – es gibt dann nur einen Typus der Äquivalenzbeziehung.²⁰

Die assoziative Beziehung wird oftmals ex negativo definiert als eine Beziehung, die weder hierarchisch noch eine Äquivalenzbeziehung ist.²¹ Hiervon kann es naturgemäß eine Vielzahl geben, klassisch wird jedoch auch dieser Beziehungstyp nicht weiter spezifiziert.

2.2 Erweiterte Relationeninventare

Bedenkt man, dass die klassischen Relationstypen in Dokumentationssprachen also maximal fünf sind (die generische, die partitive, die Instanzrelation, die Äquivalenzrelation sowie die assoziative), in der Regel sogar nur drei (die nicht näher definierte hierarchische im Allgemeinen, die Äquivalenzrelation sowie die assoziative), liegt der

¹⁹Vgl. [NISO 2005], S. 10.

²⁰Diese Definition der Äquivalenzrelation – und damit einhergehend die Einführung des Wortes „Vorzugsbegriff“ – weicht bewusst von der Definition dieser Relation in der Literatur ab. Wersig unterscheidet zwischen Begriffen und Objekten als geistige Einheiten oder Gegenstände einerseits und Bezeichnungen andererseits, die der Repräsentation von Begriffen und Objekten dienen. (Vgl. [Wersig 1978], S. 13. Natürlichsprachliche Bezeichnungen heißen hierbei Benennungen.) Die Äquivalenzrelation diene demnach dazu, Benennungen, die den gleichen Begriff repräsentieren und also als synonym zu betrachten sind, zusammenzuführen (vgl. [Wersig 1978], S. 52). Echte Synonymie im sprachwissenschaftlichen Sinne kommt jedoch nur sehr selten vor (vgl. sprachwissenschaftliche Literatur, etwa [Meibauer 2002], S. 181f.), auch Wersig plädiert dafür, in manchen Fällen begrifflicher Überschneidung Äquivalenz zu postulieren, wenn es sinnvoll erscheint, auch wenn diese nicht gegeben ist (vgl. [Wersig 1978], S. 50–52). Daher geht der Autor dieser Arbeit davon aus, dass unterschiedliche Benennungen in der Regel auch unterschiedliche Begriffe bezeichnen. Das macht, pragmatisch einen Thesaurus betrachtend – und die Entwicklung von Dokumentationssprachen soll immer pragmatisch erfolgen –, die Unterscheidung von Begriffen und Bezeichnungen hinfällig. So wird in dieser Arbeit nicht zwischen diesen unterschieden, sondern es wird davon gesprochen, dass auch die Äquivalenzrelation eine Beziehung zwischen unterschiedlichen Begriffen anzeigt, die jedoch im Kontext der jeweiligen Dokumentationssprache als gleich angesehen werden können. Ob Letzteres so ist, hängt naturgemäß von den Anwendungskontexten der Dokumentationssprache ab und kann in zwei verschiedenen Sprachen unterschiedlich entschieden werden. Der Begriff, dem gegebenenfalls der Vorrang gegeben wird, ist demnach der Vorzugsbegriff, nicht die in [Wersig 1978] auf S. 52–54 so genannte Vorzugsbenennung.

²¹Vgl. etwa [Aitchison 2000], S. 60.

Wunsch nahe, dieses Relationeninventar zu erweitern.

Ein erweitertes Inventar bedeutet naturgemäß einen größeren Aufwand in der Erschließung, bringt aber höher zu wertende Vorteile im Retrieval. Diese zeigen sich insbesondere in der Navigation im Vokabular der Dokumentationssprache, etwa innerhalb einer graphischen Oberfläche. Sind nur die genannten drei Relationstypen vorgesehen, können beispielsweise alle Unterbegriffe eines Begriffs bestenfalls alphabetisch sortiert wiedergegeben werden, obgleich sich die Unterordnung ja in der Regel aus verschiedenen Aspekten ergibt, die jedoch nicht formal in der Dokumentationssprache dargelegt sind. Dies führt insbesondere bei einer großen Zahl von Unterbegriffen, wie sie beispielsweise im Kategoriensystem der Wikipedia oft vorliegt, zu Unübersichtlichkeit. Ließen sich die Unterbegriffe aber nach den Aspekten ihrer Unterordnung – mindestens dem generischen, dem partitiven sowie dem Instanzaspekt, bestenfalls auch darüber hinaus weiter differenzierten – gruppieren, führte dies zu einem stark erhöhten Grad an Übersichtlichkeit und damit zu einer starken Verbesserung der Möglichkeiten der Navigation im Vokabular sowie der Modifikation und Erweiterung von Suchanfragen, etwa durch der Einbeziehung von Unterbegriffen in eine Suche lediglich nach einem bestimmten Aspekt. Das Gleiche gilt insbesondere für die assoziativen Relationen.

In der Vergangenheit gab es verschiedene Versuche, Dokumentationssprachen mit einer höheren Zahl an Relationstypen zu entwickeln und erweiterte Relationeninventare zur Verfügung zu stellen. [Boteram 2008] enthält auf S. 26–28 sowie S. 53f. Übersichten über diesbezügliche Literatur.

Aus den Tatsachen, dass das im Verlauf dieser Arbeit zu modellierende Relationeninventar für gemeinschaftliche Relationierungsarbeit insbesondere durch nicht informationswissenschaftlich vorgebildete Fachleute geschaffen werden soll und dass das zu entwickelnde Begriffssystem ebenfalls durch jedermann benutzbar sein soll, ergibt sich jedoch die Notwendigkeit, das Inventar möglichst allgemeinverständlich und einfach handhabbar zu gestalten. Dies kann insbesondere dadurch realisiert werden, neue und klassische Relationstypen nicht ungeordnet nebeneinander zu stellen, sondern die neuen als Verfeinerungen der drei klassischen Grundtypen zu betrachten. Diese Möglichkeit ergibt sich aus der Tatsache, dass die drei Grundtypen ihrerseits bereits jede mögliche Relation umfassen: Jede Beziehung zwischen zwei Begriffen ist entweder hierarchisch, eine Äquivalenzbeziehung oder nichts von beidem und damit assoziativ. Es liegt also nahe, jeden neuen Relationstypen in eine der drei Gruppen einzuteilen. Daraus ergibt sich die Möglichkeit, das neue Relationeninventar seinerseits hierarchisch zu gestalten. In dieser Vorgehensweise liegt nicht zuletzt ein Vorteil betreffend die Handhabung: Wer bereits mit einem Thesaurus gearbeitet hat, dem sind die drei klassischen Relationstypen bekannt. Neue Relationstypen als Verfeinerung der drei bekannten zu begreifen, fällt also leichter, als sie als neben den drei bekannten stehend anzusehen.

Ausgehend von den dargestellten klassischen Relationstypen, ergibt sich zunächst folgende Hierarchie für das klassische Relationeninventar:

1. hierarchische Relation
 - a) generische Relation
 - b) partitive Relation
 - c) Instanzrelation
2. Äquivalenzrelation
3. assoziative Relation

Jede neue Relation ist also als Verfeinerung einer dieser Relationstypen zu betrachten.²²

Den Ansatz der hierarchischen Ordnung von Relationstypen verfolgen auch Michel und Boteram, von ihnen entwickelte Inventare sind in [Michel 1997] beziehungsweise in [Boteram 2008], S. 66–73, dargestellt. Der Entwicklungsansatz von Boteram, nämlich die Schaffung eines Relationeninventars aus einem bestehenden Begriffssystem heraus, mithin also die Typisierung bereits vorhandener Relationen, wird in dieser Arbeit übernommen. Das genaue Vorgehen wird in Abschnitt 4 beschrieben.

²²Dabei sei darauf hingewiesen, dass es vor der Entwicklung des Inventars noch unklar ist, ob es neben der generischen, der partitiven und der Instanzrelation noch weitere Typen von hierarchischen Relationen geben kann. Es muss also im Ausgang der Entwicklung offengehalten werden, ob die neuen Relationen nur unterhalb von 1a, 1b, 1c sowie 2 und 3, oder auch direkt unter 1 eingeordnet werden können.

3 Das Kategoriensystem der Online-Enzyklopädie Wikipedia

Bevor die Entwicklung des neuen Relationeninventars stattfinden kann, müssen zunächst die Natur des Kategoriensystems der Wikipedia und sein derzeitiger Entwicklungsstand untersucht werden. Dies geschieht in diesem Abschnitt der Arbeit. Dabei beschränken sich die Untersuchungen des derzeitigen Entwicklungsstands auf die deutschsprachige Wikipedia.

3.1 Das Wiki-Prinzip und die Geschichte der Wikipedia

In ihren Grundzügen wurde die Wikipedia bereits in der Einleitung dieser Arbeit vorgestellt, ebenso wurden dort Aussagen über derzeit erreichte Größen der einzelnen Sprachfassungen gemacht. Im Folgenden werden die Geschichte der Wikipedia sowie das ihr zugrunde liegende Konzept knapp dargestellt. Sämtliche Angaben entstammen dem Abschnitt 2 von [Voß 2005] (S. 10 – 26), einer informatrischen Analyse der Wikipedia. Für eine ausführlichere Darstellung sei auf jene Arbeit verwiesen.

Die Wikipedia basiert auf dem Prinzip eines Wikis.²³ Ein Wiki ist ein System von Webseiten, die von jedem Nutzer direkt im Browser geändert werden können. Hierzu steht in der Regel auf jeder Seite ein Link zur Verfügung, der zu einer Bearbeitungsoberfläche führt. Dort können der Text modifiziert sowie mittels einer einfachen Auszeichnungssprache etwa Strukturierungen des Seiteninhalts vorgenommen oder Hyperlinks zu anderen Seiten ausgezeichnet werden. Ein Wiki ist nach dem Hypertext-Prinzip organisiert, es erfolgt in hohem Maße eine Verlinkung der Seiten untereinander. Im Gegensatz zu Content-Management-Systemen gibt es kein Konzept von Benutzerrechten, vielmehr sind Änderungen durch jeden Betrachter, oftmals sogar ohne vorherige Anmeldung möglich. Um Änderungen nachvollziehbar zu machen, bleibt jede Fassung jeder Seite gespeichert. Es ist also möglich, jede Fassung im Nachhinein erneut zu betrachten beziehungsweise die Unterschiede zur jeweils aktuellen Fassung anzeigen zu lassen. Das erste Wiki-System wurde 1995 entwickelt.

Die Begründung der Wikipedia erfolgte Anfang 2001 aus der Online-Enzyklopädie Nupedia heraus, die von zwei Privatpersonen, Jimmy Wales und Larry Sanger, gegründet worden war. Es handelte sich dabei zunächst um eine Erweiterung von Nupedia, deren Artikel zuvor von Fachautoren geschrieben wurden und einem Peer Review unterlagen, um ein Wiki, das für Vorarbeiten an Artikeln gedacht war. Dieses Wiki entwickelte jedoch eine hohe Eigendynamik – nach einem Monat gab es mehr als 1.000, Anfang September 2001 mehr als 10.000 Artikel –, was im weiteren Verlauf zur Aufgabe des ursprünglichen Konzeptes der Nupedia führte. Die Online-

²³Das Wiki-Prinzip wird auf S. 10f. von [Voß 2005] erläutert. Auf jene Darstellung stützt sich dieser Absatz.

Enzyklopädie wurde nunmehr als Wikipedia genanntes Wiki fortgeführt. Im Juni 2003 erfolgte die Gründung der „Wikimedia Foundation“, die seitdem deren Server betreibt sowie die Namensrechte am Begriff „Wikipedia“ hält. Die Mitarbeit an der Wikipedia ist für jeden Internetnutzer möglich, entweder unter einem Benutzernamen nach vorheriger Anmeldung, die jedoch keinen Beschränkungen unterliegt, oder ganz ohne Anmeldung, wobei die vorgenommenen Änderungen im Zusammenhang mit der IP-Adresse, also der Internet-technischen Adresse des Rechners, von dem sie getätigt wurden, gespeichert werden. Neben der eigentlichen Artikelseite, die den Artikeltext enthält, gibt es zu jedem Lemma eine Diskussionsseite, mittels der über den Artikeltext diskutiert werden kann, ohne dass dieser immer direkt geändert werden muss. Die Inhalte der Wikipedia unterliegen der Creative-Commons-Lizenz „CC-by-sa-3.0“ („Creative Commons Attribution-ShareAlike 3.0 Unported“), die die freie Zurverfügungstellung und Verwendung ermöglicht. Die der Wikipedia zugrunde liegende Software heißt MediaWiki.

Das Wiki-Prinzip hat aus Nutzersicht naturgemäß den Vorteil, Inhalte rasch verändern zu können, ohne Änderungswünsche bei einer zentralen Stelle einreichen zu müssen. Dagegen wurden jedoch Einwände vorgebracht, die prinzipbedingte Nachteile herausstellen. [Weller 2006] benennt diese in einem Satz: „Andernfalls hätte das Ontologie-Wiki [jene Arbeit beschäftigt sich mit der kooperativen Ontologieerstellung, A. M.] mit ähnlichen Problemen zu kämpfen, die auch bei anderen offenen Wikis, wie vor allem Wikipedia, regelmäßig kritisiert werden, wie z. B. mangelnde Kontrolle über die Einträge, fehlende Zuverlässigkeit und möglicher Vandalismus.“ (S. 231.)²⁴ Bei der Wikipedia stellt sich zudem die Frage nach der Motivation der Autoren: Gibt es überhaupt genug Personen, die motiviert sind, ehrenamtlich an einer Enzyklopädie zu arbeiten?

Die Wikipedia versucht, den genannten Nachteilen durch umfangreiche Verhaltenskodizes und Richtlinien für Autoren, insbesondere in Bezug auf das Wahren eines neutralen Standpunktes beim Schreiben, entgegenzuwirken. Viele Autoren beobachten durchgehend die Entwicklung von Artikeln, werden so auf fehlerhafte Darstellungen rasch aufmerksam und können diese korrigieren. Daneben gibt es eine Gruppe von Nutzern, genannt Administratoren, mit Sonderrechten. Eine Analyse der Wikipedia hinsichtlich der Vor- und Nachteile sowie eine Untersuchung der Motivation der Mitarbeitenden ist nicht Teil dieser Arbeit und soll an dieser Stelle auch nicht erfolgen. Hingewiesen sei jedoch auf die genannten Umfänge einzelner Sprachfassungen der Wikipedia²⁵ sowie auf zwei Arbeiten: [Reagle 2004] benennt praktische Probleme in der gemeinschaftlichen Arbeit innerhalb der Wikipedia und stellt dar, wie diese gelöst werden. [Frost 2006] ist eine umfangreiche Untersuchung der deutschsprachigen Wikipedia aus zivilgesellschaftlicher Perspektive, die unter an-

²⁴Vgl. zu dieser Thematik auch [Frost 2006], S. 9.

²⁵Vgl. S. 7 dieser Arbeit.

derem Vergleiche mit anderen Formen bürgerschaftlichen Engagements zieht. Nach dem Eindruck des Autors dieser vorliegenden Arbeit existiert in der Wikipedia heute ein funktionierendes System der sozialen Kontrolle, das dazu führt, dass der allergrößte Teil der Inhalte sinnvoll ist, von einem neutralen Standpunkt ausgeht, der Wahrheit entspricht sowie einer Enzyklopädie angemessen ist. Ebenso wird Vandalismus in der Regel rasch erkannt und, da jede Fassung jedes Artikels gespeichert ist, rückgängig gemacht.

3.2 Entstehung und Entwicklungsstand des Kategoriensystems

Eine Einführung in die Entstehung des Kategoriensystems gibt [Voß 2006], eine Untersuchung des Systems aus dem Jahr 2006, auf S. 1: Laut Voß gab es schon seit der Entstehung der Wikipedia Bemühungen um eine Sortierung der Artikel, zunächst in Form von Listen sowie Listen von Listen. Ab Mai 2004 war mit der Einführung einer neuen Version der Software MediaWiki die Möglichkeit der Kategorisierung von Artikeln und Kategorien selbst gegeben.

Mit einer Kategorie, in die ein Artikel oder eine andere Kategorie eingeordnet werden können, ist in diesem Zusammenhang in der Regel eine natürlichsprachliche Bezeichnung gemeint, die einen Begriff darstellt, der mit der eingeordneten Sache in irgendeinem Zusammenhang steht.²⁶ Durch die Einordnung in eine Kategorie wird also ein begrifflicher, semantischer Zusammenhang zwischen dem durch den Kategoriennamen Bezeichneten und dem durch den Namen des Eingeordneten Bezeichneten hergestellt. Der Prozess dieser Einordnung ist zum einen, betrachtet man das Einzuordnende als etwas zu Erschließendes, mit der Vergabe eines Schlagwortes in einem dokumentarischen Zusammenhang vergleichbar, zum anderen ist er, betrachtet man die Namen der Kategorie und des Einzuordnenden als Begriffe, mit der Herstellung einer Relation zwischen zwei Begriffen in einem Begriffssystem vergleichbar. Es ist darauf hinzuweisen, dass der Begriff der Kategorie hier also *nicht* in seiner philosophischen Bedeutung einer grundlegenden Denk- oder Seinsform verstanden wird, sondern vielmehr das darstellt, was im Dokumentationswesen ein Begriff oder ein Schlagwort ist. Um Missverständnisse in Bezug auf die Wikipedia zu vermeiden, wird der Begriff der Kategorie im Verlauf dieser Arbeit lediglich im Sinne der Wikipedia verwendet.

Unklarheit herrschte unter den an der Wikipedia Mitwirkenden bereits seit dessen Einführung über den Sinn des Kategoriensystems. Eine die Problemstellungen gut

²⁶Daneben gibt es auch Kategorien, die der Aufnahme von Artikeln nach formalen Kriterien dienen, beispielsweise der Aufnahme von Artikeln, die Listen enthalten. Diese sind jedoch quantitativ wenige. Außerdem existieren die in Fußnote 32 auf S. 20 beschriebenen sogenannten Wartungskategorien, die jedoch ebenfalls nur einen geringen Teil ausmachen.

zusammenfassende Kritik von Voß in der Mailingliste *wikipedia-l* ([Voß 2004a]) bezog sich insbesondere auf die Fragen, was mit der Kategorisierung überhaupt erreicht werden und wie diese ausgeprägt sein soll – „Are we trying to index articles with subject heading [sic!], using a thesaurus, a classification or even a structure ontology?“ –, ebenso wurde darauf hingewiesen, dass Inhaltserschließung ein komplexes Thema sei²⁷, womit die Frage angedeutet wird, ob ein Vorhaben in diese Richtung also überhaupt in die Hände der Allgemeinheit übergeben werden kann. Daneben ergäben sich Probleme durch technische Einschränkungen, etwa durch die Tatsache, dass keine sogenannten Weiterleitungen auf Kategoriennamen möglich sind, also keine Synonyme von Kategorienbezeichnungen gespeichert werden können.²⁸ [Voß 2004a] kommt zu folgendem Schluss:

„Given the reasons I strongly recommend to stop using the categories and to focus on writing and improving good articles. [...]

Indeed classifying Wikipedia articles is very interesting and will become more important, but this should be an independent project – maybe in a ‚Classifipedia‘ or ‚Categorypedia‘ that links to Wikipedia articles.

You know—librarians normally do not write the books they organize and search engine experts do not write the websites they crawl, so let’s focus on what we can do the best: creating the most detailed, most understandable and freest encyclopedia in the history of mankind!“

Dennoch wurde die Möglichkeit, Kategorien zu schaffen, quantitativ betrachtet gut angenommen.²⁹ Es entstand in der deutschsprachigen Wikipedia ein dichtes Netz von Kategorien, in das fast jeder Artikel eingebunden ist. Wie bereits in der Einleitung dieser Arbeit dargelegt, ist der einzige im Kategoriensystem technisch ausgewiesene Relationstyp der hierarchische. Eine solche Relation entsteht eben durch die Einordnung eines Artikels oder einer Kategorie in eine andere Kategorie. Letztere wird dadurch zur direkt übergeordneten Kategorie der eingeordneten Sache. Durch die Einordnung dieser in eine weitere Kategorie und einer Einordnung dieser weiteren Kategorie in eine nächste ergibt sich eine mehrstufige Hierarchie. Dadurch, dass in der Praxis viele Artikel und Kategorien mehrere direkt übergeordnete Kategorien haben, entsteht eine stark polyhierarchische Struktur, an deren Enden jeweils die Artikel stehen, die naturgemäß keinerlei Untergeordnetes haben können.

²⁷„It’s not easy to know how to do it in [sic!] the right way because subject indexing is a complex issue [...]. ([Voß 2004a].)

²⁸Siehe hierzu ebenfalls [Voß 2004a] sowie [Voß 2004b]. Eine sogenannte Weiterleitung ist auf der Ebene von Artikeln möglich. Sie bedeutet, dass neben dem Namen eines Artikels ein weiterer Begriff gespeichert wird, der als synonym zum Artikelnamen angesehen wird. Ruft ein Nutzer diesen anderen Begriff auf, so wird er zum eigentlichen Namen und damit zum eigentlichen Artikel weitergeleitet. Das Einrichten einer solchen Weiterleitung kommt also der Herstellung einer Äquivalenzbeziehung zwischen dem Artikelnamen und dem Begriff, von dem weitergeleitet wird, gleich. Bei Kategorien sind Weiterleitungen, wie beschrieben, jedoch nicht möglich.

²⁹Vgl. die in der Einleitung dieser Arbeit auf S. 7 genannten Zahlen.

Ebenso wie für Artikel existieren sogenannte Namenskonventionen, also Regeln zur Benennung, auch für Kategorien.³⁰ Bei Kategorien wird bei entsprechender Notwendigkeit die Verwendung von Qualifikatoren vorgeschrieben, die in Klammern an die Bezeichnung angehängt werden. Diese haben wie in anderen kontrollierten Vokabularen die Funktionen der Differenzierung von Homonymen. Zu jeder Kategorie kann zudem eine Beschreibung in Textform gespeichert werden. Von dieser Möglichkeit wird derzeit jedoch nicht immer Gebrauch gemacht. Wie bereits erwähnt, sind, im Gegensatz zu Artikeln, keine Weiterleitungen auf Kategoriennamen möglich. Es können also keine Synonyme zu Kategorienbezeichnungen gespeichert werden.

Ruft man eine Kategorie innerhalb der Wikipedia mit einem Browser auf, so erscheinen, wenn vorhanden, deren Beschreibung und darunter, in Listenform und mit Hyperlinks versehen, die direkt untergeordneten Kategorien und Artikel. Bei den untergeordneten Kategorien gibt es die Möglichkeit, durch einen Klick auf ein danebenstehendes Symbol die jeweils weiteren auf den nächstunteren Stufen – nicht jedoch darunterliegende Artikel – anzuzeigen. Die direkt übergeordneten Kategorien stehen, mit Hyperlinks versehen, immer am Ende einer Kategorien- oder Artikelseite. Mehr als eine übergeordnete Hierarchiestufe ist in der Weboberfläche der Wikipedia nicht auf einen Blick sichtbar. Mit dem Programm Catgraph ([Catgraph 2009]) steht jedoch ein Graphviz benutzendes Werkzeug zur Verfügung³¹, das, ausgehend von einer Kategorie oder einem Artikel die Relationen zu benachbarten Kategorien und darüber hinaus als Graph darstellen kann. Zwei Beispiele bieten die Abbildungen 1 und 2.

Innerhalb der Wikipedia werden die Kategorien als Mittel der inhaltlichen Erschließung der Artikel gedacht und eingesetzt. So enthalten die Richtlinien für Wikipedia-Autoren in Bezug auf Kategorien den folgenden Satz: „Kategorien sind in der Wikipedia ein Mittel, mit dem Seiten nach bestimmten Merkmalen eingeordnet werden können.“ ([Wikipedia 2010e].) Die im vorigen Absatz beschriebene Struktur wird hier also ohne die Artikel, ohne die Enden der von Catgraph erzeugten Graphen gedacht, vielmehr dienen die Kategorien, informationswissenschaftlich gesprochen, der Verschlagwortung der Artikel. Dabei ist die Zahl der Schlagwörter pro Artikel nicht festgelegt. Der Prozess der Zuordnung von Artikeln zu Kategorien ist also vergleichbar mit dem Prozess der Zuordnung von kontrollierten Schlagwörtern beziehungsweise Deskriptoren eines Thesaurus zu einer Dokumentationseinheit.³²

³⁰Die Namenskonventionen in der deutschsprachigen Wikipedia sind auf der Seite [Wikipedia 2010d] beziehungsweise den darunterliegenden zusammengestellt.

³¹Zu Graphviz, das auch in dieser Arbeit zur Visualisierung der neuen Relationenstrukturen eingesetzt wird, vgl. Abschnitt 5.3.5.

³²Neben der inhaltlichen Einordnung der Artikel zum Zweck des späteren Auffindens wird das Kategoriensystem auch für statistische Analysen der Wikipedia eingesetzt. Vgl. hierzu ebenfalls [Wikipedia 2010e]. Außerdem gibt es einige Kategorien, die nicht den Inhalt eines Artikels beschreiben, sondern der Wartung dienen, etwa um Artikel zu markieren, die der Überarbeitung bedürfen. Zum Beispiel können Artikel, deren Neutralität umstritten ist, in die Wartungskate-

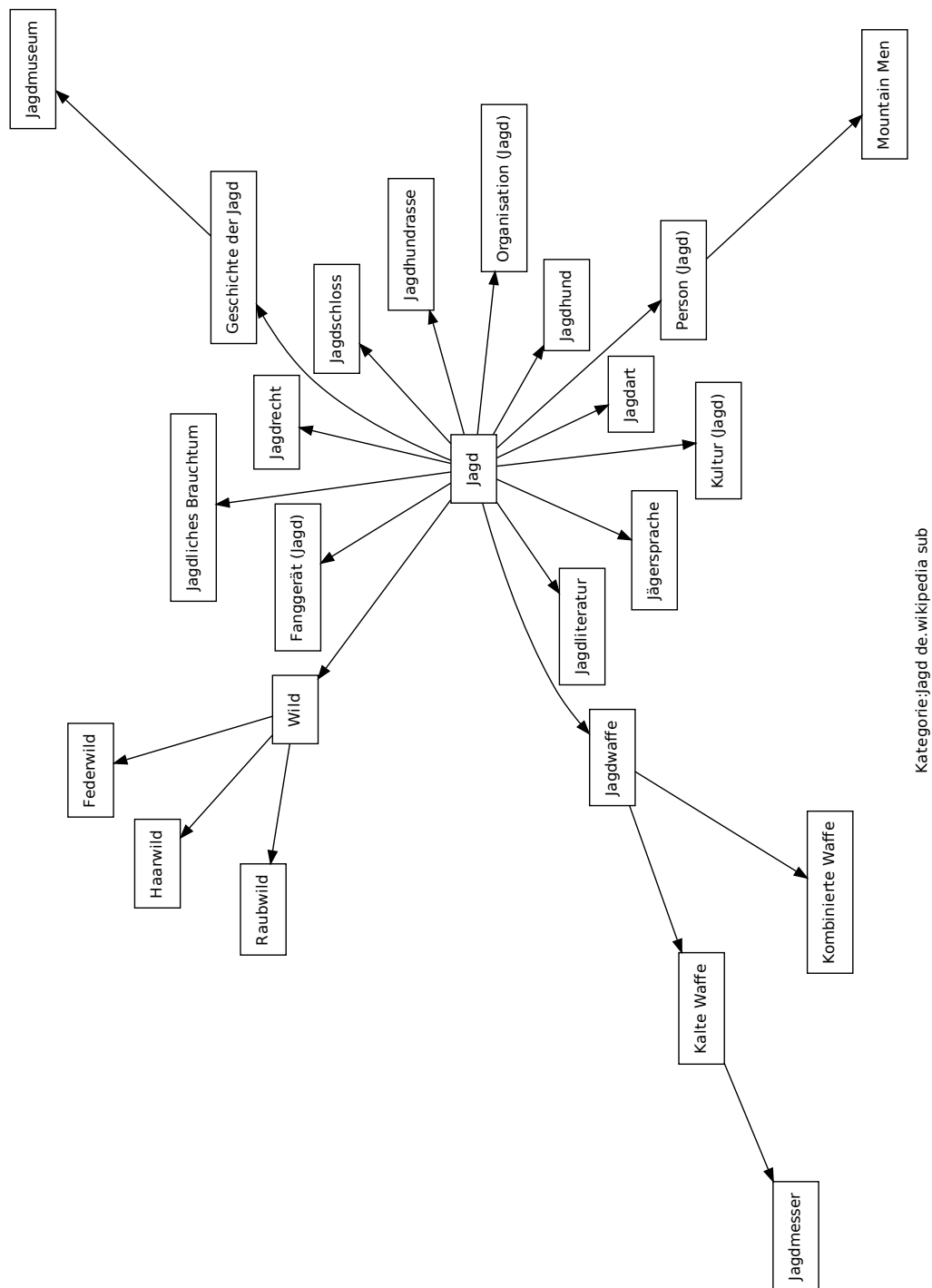


Abbildung 1: Der Kategorie „Jagd“ untergeordnete Kategorien – nicht Artikel – im Kategoriensystem der deutschsprachigen Wikipedia (Darstellung mit Catgraph als Wolke)

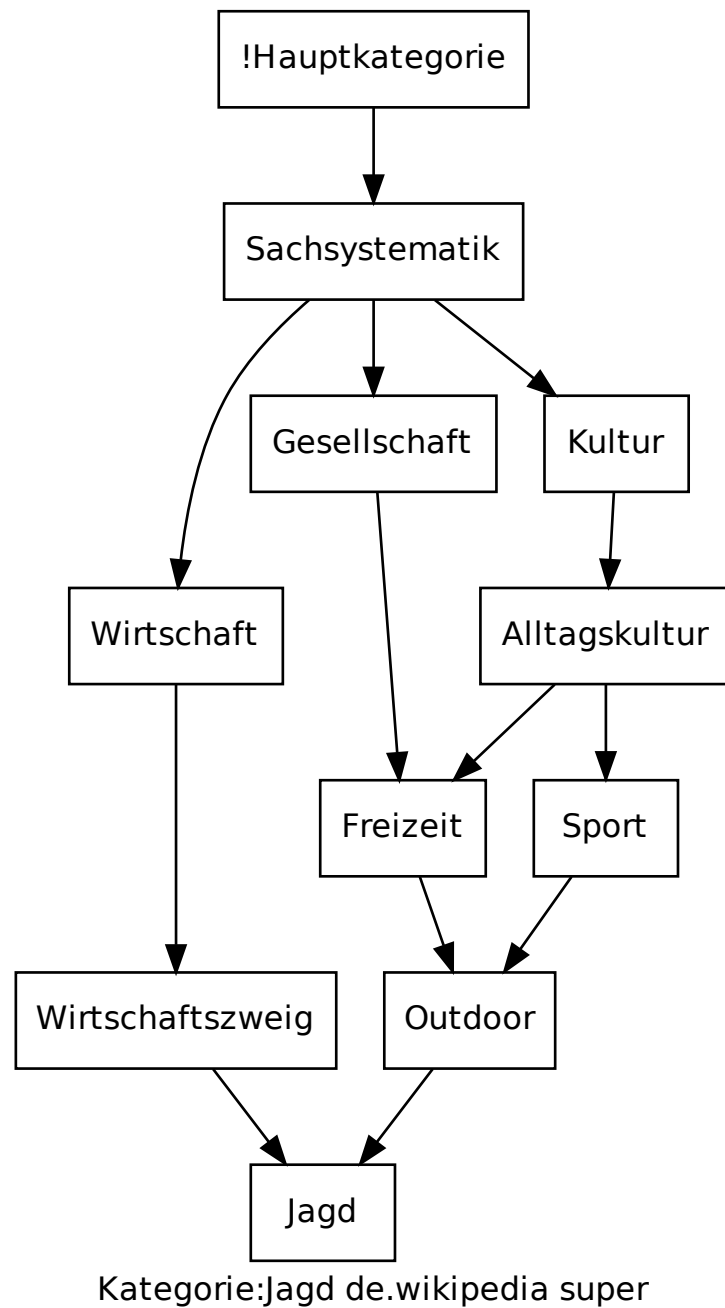


Abbildung 2: Der Kategorie „Jagd“ übergeordnete Kategorien im Kategoriensystem der deutschsprachigen Wikipedia (Darstellung mit Catgraph als Baum)

Dennoch werden bei Betrachtungen der Begriffsbeziehungen innerhalb des Kategoriensystems oftmals die Artikel wieder in die Struktur eingefügt und nicht als das zu Verschlagwortende betrachtet, sondern ihre Namen als Teil eines Begriffssystems, das also aus den Namen der Kategorien und den Namen der Artikeln besteht, die als Begriffe gedacht werden. So tut es auch [Voß 2006] auf S. 3 beim Vergleich der übergeordneten Relationen des Deskriptors „stomach neoplasms“ aus den Medical Subject Headings (MeSH) und des Artikels „stomach cancer“ aus der englischsprachigen Wikipedia. Eine solche Vorstellung eines Begriffssystems aus Kategorien- und Artikelnamen, also der Einbeziehung der Relationen zwischen Kategorien- und Artikelnamen in das System, liegt auch dieser Arbeit zugrunde. Sie kommt der oben zitierten Begriffe einer „Classifipedia“ oder „Categorypedia“ nahe, wenn auch nicht als eigenständiges Projekt, sondern innerhalb der Wikipedia.

Am Aufbau des Kategoriensystems kann sich entsprechend der Natur der Wikipedia jeder Internetnutzer beteiligen, schlicht bereits indem er etwa einem Artikel eine Kategorie zuweist oder Änderungen innerhalb der Kategorienstruktur vornimmt. Daneben haben sich in der deutschsprachigen Wikipedia jedoch Strukturen entwickelt, die den Aufbau des Kategoriensystems leiten sollen. Dabei wurde die Pflege der Kategorien zu vielen Gegenstandsbereichen von sogenannten Fachbereichen übernommen – Benutzern, die sich in sogenannten Portalen, Redaktionen oder WikiProjekten zusammengetan haben, um die Darstellung eines Gegenstandsbereichs innerhalb der deutschsprachigen Wikipedia voranzutreiben, wozu auch der entsprechende Teil des Kategoriensystems gehört.³³ Für Kategorien, die nicht Teil eines Fachbereichs sind, existiert das WikiProjekt „Kategorien“.³⁴

Erkennbar ist die Tendenz, die Kategorien der Fachbereiche nicht nebeneinander zu stellen, sondern sämtliche Kategorien unter einer einzigen Hauptkategorie zusammenzuführen. Diese trägt den Namen „!Hauptkategorie“.

3.3 Typologie des Kategoriensystems als Begriffssystem und derzeitige Problemstellungen

Wie im vorigen Abschnitt dargelegt, war bei der Einführung des Kategoriensystems unklar, wozu dieses dienen und wie es ausgestaltet werden soll. Diese Unklarheiten sind bis heute nicht vollständig ausgeräumt, denn es existieren keine allgemein anerkannten Richtlinien, die die Gestaltung des gesamten Kategoriensystems an sich

gorie „Wikipedia: Neutralität“ eingeordnet werden. So wird es möglich, alle Artikel, auf die jenes zutrifft, zusammen zu betrachten und gegebenenfalls zu überarbeiten. Die Wartungskategorien machen jedoch im gesamten Kategoriensystem einen quantitativ sehr kleinen Teil aus.

³³Eine Liste der Fachbereiche enthält [Wikipedia 2010f]. Einen Überblick über Portale, Redaktionen und WikiProjekte liefern das sogenannte Autorenportal [Wikipedia 2010g] sowie die dort verlinkten Seiten unter der Überschrift „Initiativen und Projekte“.

³⁴Vgl. [Wikipedia 2010h].

betreffen.³⁵ Voß fasste bereits im Juni 2004 zeitlich nah zu der im vorigen Abschnitt zitierten Mail zusammen:

„Sowohl über Ziele und Zwecke der Verwendung von Kategorien innerhalb der Wikipedia als auch über die dabei empfehlenswerte oder richtige Vorgehensweise herrscht bislang keine Einigkeit. Konsens herrscht höchstens darüber, dass die Kategorien zu irgend einer Art von Inhaltserschließung verwendet werden sollen. Die Diskussion ist innerhalb von wenigen Wochen praktisch unüberschaubaren Flut von Vorschlägen, Meinungen und Problembeschreibungen angeschwollen. [sic!] [...]

Konkret lassen sich folgende Ansätze feststellen:

- (Teilkontrollierte) Verschlagwortung
- Thesaurus-artige Erschließung
- Einordnung in eine feste Systematik (monohierarchisch oder multihierarchisch, Analytische Klassifikation oder Facettenklassifikation ...)
- Bildung einer logischen Ontologie mit starker Vererbung

Eine systematische Reflexion und Einordnung der Ansätze findet aufgrund der fehlenden dokumentarischen Grundkenntnisse nur in Ansätzen statt. Stattdessen werden mehrheitlich Meinungskundgebungen und Einzelbeispiele und -vorschläge diskutiert oder untereinander gestellt.“ ([Voß 2004b])

Ogleich durch die bislang stattgefundenen Kategorisierungsprozesse pragmatisch Fakten geschaffen wurden, ist das grundlegende Ordnungssystem bis heute Gegenstand von Diskussionen, sodass die Kritik von Voß nach wie vor als zumindest eingeschränkt zutreffend bezeichnet werden muss. Das WikiProjekt „Kategorien“ fasst heute zusammen:

„Verschiedene grundlegende Systeme [der Kategorienbildung] wurden diskutiert, die heute praktizierte Lösung ist aber eine Sammlung von Teilstrukturen, die unabhängig in den jeweiligen Fachbereichen entwickelt werden.“ ([Wikipedia 2010h])

Für die Kategorisierung in der deutschsprachigen Wikipedia ergibt sich also zunächst das Problem, dass kein verbindliches umfangreiches Ordnungskonzept existiert, das eine wahrheitsgemäße Aussage darüber einschließt, welchem Typ eines Begriffssystems beziehungsweise einer Dokumentationssprache überhaupt gefolgt werden soll. Lediglich gibt es innerhalb der Richtlinien für Wikipedia eine kurze Anleitung zum Kategorisieren, die sich jedoch hauptsächlich mit der Benennung und

³⁵Die archivierten Diskussionen über das Kategoriensystem in seiner Gesamtheit sowie über einzelne Kategorien, beginnend auf der Seite [Wikipedia 2010i], sind sehr umfangreich.

der Größe von Kategorien, nicht mit deren Struktur, befasst. Diese Anleitung nennt das Kategoriensystem einen „Thesaurus, der teilweise nach Art einer Klassifikation aufgebaut ist und teilweise Qualifikatoren benutzt.“ ([Wikipedia 2010e]) In seiner Untersuchung des Kategoriensystems schreibt Voß: „Namely the category system is a thesaurus with hierarchical relationships [...]“ ([Voß 2006])

Die so ins Spiel gebrachten Bezeichnungen von Dokumentationssprachen sollen in Bezug auf das Kategoriensystem im Folgenden näher untersucht werden.

Die zugrunde liegende Software ermöglicht – dies ist ein bereits früh von Voß dargelegter Kritikpunkt³⁶, der bis heute gültig ist – im Kategoriensystem der Wikipedia nur einen Relationstypen: den, der als hierarchisch gilt. Es kann jedoch nicht davon ausgegangen werden, dass jede ausgewiesene Relation wirklich eine hierarchische im Sinne der drei in Abschnitt 2.1 beschriebenen klassischen Hierarchierelationen ist. Vielmehr sind darunter auch assoziative Relationen, die in Ermangelung einer Alternative technisch als hierarchisch gespeichert wurden.³⁷ Einzig das Vorhandensein von Äquivalenzbeziehungen – hiermit sind nicht die Weiterleitungen auf Artekelebene gemeint, sondern Beziehungen zwischen Kategorien selbst – erscheint unwahrscheinlich, da es keinen Sinn hätte, eine Kategorie zu erschaffen, die das Gleiche beschreibt wie eine andere, und erstere dann letzterer zuzuordnen.³⁸

Betrachtet man die Artikel, wie es die Richtlinien für Wikipedia-Autoren tun, nicht als Teil des Begriffssystems, sondern als das zu Erschließende, erzeugt die Tatsache, dass jeder Artikel einer beliebigen Zahl von natürlichsprachlich benannten Kategorien zugeordnet wird, eine Nähe zum Thesauruskonzept. Da jedoch grundlegendes Merkmal von Thesauri die terminologische Kontrolle im Sinne der Ausweisung von Vorzugsbegriffen und Synonymen ist³⁹, erscheint es problematisch, ein Begriffssystem, das, lässt man die Artekelebene außen vor, in Gänze ohne Äquivalenzbeziehungen auskommt, als Thesaurus zu bezeichnen. Zumindest wäre ein Thesaurus ohne jegliche Äquivalenzbeziehung in hohem Maße unvollständig. Ebenso erfordert ein Thesaurus die einheitliche Formulierung von Deskriptoren. Zwar kennt die Wikipedia Richtlinien zur Benennung von Kategorien und Artikeln⁴⁰, doch kann aufgrund ihrer auf gemeinschaftliche Arbeit gestützten Natur nicht davon ausgegangen

³⁶Vgl. den vorigen Abschnitt sowie [Voß 2004a] und [Voß 2004b].

³⁷Vgl. hierzu die im Rahmen dieser Arbeit nur anhand der bestehenden Relationen entwickelten Relationeninventare, die eine Vielzahl assoziativer Relationen enthalten, sowie die entsprechenden neu relationierten Vokabulare auf der beiliegenden CD-ROM. Dennoch benutzt diese Arbeit im Folgenden aus sprachlichen Gründen weiterhin die Begriffe der ober- und der untergeordneten Kategorie, da diese ja von der der Wikipedia zugrunde liegenden Software so gedacht sind.

³⁸Möglich ist allenfalls, dass eine Kategorie existiert, die wie ein Artikel benannt ist. Dann sollte der betreffende Artikel naturgemäß auch der Kategorie zugeordnet sein. Technisch gesehen wäre dies eine Beziehung zwischen zwei gleichen Bezeichnungen. Diese Vorkommnisse sollten bei einer Betrachtung des Kategoriensystems im Hinblick auf die Relationen jedoch ausgeklammert werden.

³⁹Vgl. [Wersig 1978], S. 47–80.

⁴⁰Vgl. die bereits im vorigen Abschnitt erwähnten sogenannten Namenskonventionen unter [Wikipedia 2010d]

werden, dass diese Regeln überall auch in kleinsten Details eingehalten werden.⁴¹

Auch eine Bezeichnung des Kategoriensystems als Klassifikation fällt schwer.⁴² Dagegen sprechen die starke Polyhierarchisierung innerhalb der Kategorien sowie die Tatsache, dass ein Artikel zu mehreren Kategorien gehören kann, wobei die Zahl der Kategorien jedoch nicht festgelegt ist und ebenso keinen definierten Facetten folgt wie in einer Facettenklassifikation. Es kann lediglich in dem Versuch, sämtliche Kategorien unter einer Hauptkategorie anzuordnen, eine gedankliche Nähe zu Universalklassifikationen vermutet werden.

Zusammenfassend entspricht das Kategoriensystem der Wikipedia also keinem im Dokumentationswesen bekannten Typus von Begriffssystemen beziehungsweise Dokumentationssprachen, lediglich weist es Merkmale bestimmter Dokumentationsprachen auf. Festgehalten werden kann nur, dass es sich bei dem Kategoriensystem um ein Begriffssystem handelt, das nur einen Relationstypen enthält, der lediglich bedeutet, dass zwischen zwei Begriffen irgendeine in ihrer Natur nicht näher spezifizierte Beziehung besteht. Das bedeutet, dass die Darstellung des Kategoriensystems als Baumstruktur – die ja Hierarchien impliziert – irreführend ist und es sich vielmehr um eine Netzstruktur handelt, die ja der Begriff des semantischen Netzes andeutet, jedoch mit nur einem Relationstypen.⁴³

Aus dem Fehlen eines verbindlichen Konzepts betreffend die Struktur des gesamten Kategoriensystems ergibt sich, bereits im obigen Zitat des WikiProjekts „Kategorien“ benannt, eine weitere Problemstellung. Wie bereits angedeutet, zerfällt das Kategoriensystem in zwei Teile: solche, die durch Fachbereiche gepflegt werden, und solche, die außerhalb der Fachbereiche stehen und also vom WikiProjekt „Kategorien“ gepflegt werden.⁴⁴ Innerhalb eines Fachbereichs existieren üblicherweise Regeln zur Kategorisierung, die jedoch nicht immer nach dokumentarischen Gesichtspunkten niedergeschrieben sind, sondern oft ungeschrieben befolgt werden oder sich in Diskussionen ergeben haben.⁴⁵ Das führt, sofern sich ein Fachbereich dem Kategoriensystem ausführlich widmet, zu einheitlichen und sinnvollen Kategorisierungen unterhalb der obersten Kategorie des Fachbereichs.⁴⁶ Im WikiProjekt „Kategorien“ wird hierzu ausgeführt:

⁴¹In den zwei untersuchten Gegenstandsbereichen erfolgte die Benennung der Kategorien und Artikel jedoch in fast jedem Fall einheitlich und entsprechend den Namenskonventionen.

⁴²Zur Natur von Klassifikationen vgl. die Kapitel 8 bis 10 von [Bertram 2005].

⁴³Ein weiterer Relationstyp tritt erst hinzu, wenn man die Artikelweiterleitungen in das Begriffssystem einbezieht und als Äquivalenzbeziehungen betrachtet. In dieser Arbeit werden die Weiterleitungen jedoch nicht näher betrachtet.

⁴⁴Dass ein Artikel oder eine Kategorie innerhalb eines Fachbereichs angesiedelt ist, bedeutet dabei keine Ausschließlichkeit. Dieselbe Kategorie oder derselbe Artikel können zusätzlich im Teilsystem eines anderen Fachbereichs oder außerhalb eines Fachbereichs erscheinen.

⁴⁵Die niedergeschriebenen Regeln zur Kategorienstruktur eines Fachbereichs sind, sofern vorhanden, innerhalb der bereits genannten tabellarischen Übersicht der Fachbereiche [Wikipedia 2010f] verlinkt.

⁴⁶Die von einem Fachbereich gepflegten Kategorien laufen in der Regel auf eine einzige oberste Kategorie zu.

„Per Meinungsbild wurde einmal festgelegt, dass das Kategoriensystem der Wikipedia ein System der Fachkategorien ist. Daraus folgt, dass die Kategorienbäume ein für den jeweiligen Fachbereich sinnvolles, themenbezogenes System darstellen sollen. Dementsprechend hat sich die Einordnung der Unterkategorien vorrangig an den Bedürfnissen des Fachbereiches auszurichten. Diese können nicht abstrakt festgelegt werden, daher müssen die Abstimmungen darüber von den Fachleuten im jeweiligen Fachbereich getroffen werden. Die meisten Fachbereiche verfügen bereits über ein ausgefeiltes Kategoriensystem.“ ([Wikipedia 2010k])

Inkonsistenzen treten also in der Regel an zwei Stellen auf: zum einen zwischen zwei Fachbereichen, die unterschiedlichen Kategorisierungsregeln folgen, zum anderen in den Kategorien, die außerhalb der Fachbereiche liegen, denn dort existieren keine klaren Regeln. Beim Teil des Kategoriensystems außerhalb der Fachbereiche sind es vor allen Dingen die bereits genannten Versuche, sämtliche Kategorien und also sämtliche Fachbereiche unterhalb einer einzigen Hauptkategorie anzuordnen, die Probleme bereiten. Auf diese Problemstellung wird im Folgenden näher eingegangen.

Unterhalb der obersten Kategorie des Gesamtsystems, genannt „!Hauptkategorie“, befinden sich nach dem Stand vom Juni 2010 folgende direkte Unterkategorien:

- Liste
- Fachübergreifende Schlagwortkategorie
- Objektkategorie
- Räumliche Systematik
- Sachsystematik
- Zeitliche Systematik

Bei der räumlichen sowie der zeitlichen Systematik handelt es sich wahrscheinlich um Versuche der Etablierung einer Facettenklassifikation. Die Strukturen unterhalb dieser Kategorien bleiben jedoch ohne Kenntnis der ihrer Schaffung zugrunde liegenden Diskussionen unklar. Die Objektkategorie wird im WikiProjekt „Kategorien“ folgendermaßen beschrieben: „In einer Objektkategorie werden Artikel eingeordnet, die ein Objekt beschreiben.“ ([Wikipedia 2010k]) Eine Definition des Begriffs „Objekt“ erfolgt jedoch nicht. Auch die Strukturen innerhalb jener Kategorie sind nicht intuitiv verständlich. Das Gleiche gilt für die „fachübergreifende Schlagwortkategorie“. Die Kategorie „Liste“ dient der Aufnahme von Artikeln, die zum größten Teil aus Listen bestehen, und ordnet Artikel somit zunächst nicht nach inhaltlichen, sondern nach formalen Gesichtspunkten.

Unter den Kategorien, die der inhaltlichen Ordnung von Artikeln dienen, scheint einzig das Konzept der „Sachsystematik“ klar: Hierunter liegen die obersten Kategorien der meisten Fachbereiche, zum Teil als direkte Unterkategorien der Kategorie „Sachsystematik“, zum Teil auf weiteren Stufen. Dennoch sind auch hier nicht sämtliche Namen der direkten Unterkategorien intuitiv verständlich. Sie lauten derzeit: Energie, Fiktion, Geschichte, Gesellschaft, Gesundheit, Kommunikation, Kultur, Kunst, Literatur, Organisationen, Philosophie, Planen und Bauen, Recht, Religion, Sprache, Technik, Verkehrswesen, Wirtschaft, Wissenschaft sowie Zeit.

Zur weiteren Verdeutlichung der Zuordnungsprobleme, die sich durch das Fehlen eines Kategorisierungsregelwerks außerhalb der Fachbereiche – und in jenem Beispiel zum Teil auch durch mögliche übermäßige Kategorisierungen innerhalb eines Fachbereichs – ergeben, zeigt Abbildung 3 auf S. 29 die mit dem Werkzeug Catgraph dargestellte Position des Artikels „Agnus Dei“ im Kategoriensystem.⁴⁷

Zusammenfassend muss in der Betrachtung des gesamten Kategoriensystems also zwischen zwei Typen von Subsystemen differenziert werden:

1. Ein Kategoriensystem eines Fachbereichs, also sämtliche direkten und indirekten Unterkategorien der Hauptkategorie des Fachbereichs. Ein solches Teilsystem muss unabhängig vom Rest des Systems betrachtet werden. Es kann konsistent und sinnvoll sein, sofern innerhalb des Fachbereichs sinnvolle Regeln zur Kategorisierung existieren und diese auch angewandt werden.
2. Das Kategoriensystem außerhalb der Fachbereiche. Hier ergeben sich viele Unverständlichkeiten.

3.4 Anwendungen des Kategoriensystems außerhalb der Wikipedia

Die breite Themenabdeckung der Wikipedia und, damit verbunden, des Kategoriensystems, hat bereits zu Anwendungen des Systems außerhalb der Wikipedia geführt. Diese beschränken sich jedoch bislang auf das Kategoriensystem der englischsprachigen Wikipedia.

Zum einen übernimmt das Projekt DBpedia⁴⁸, das sich mit der Extraktion strukturierter Informationen aus der Wikipedia und deren Bereitstellung innerhalb einer Ontologie⁴⁹ beschäftigt, auch das Kategoriensystem der englischsprachigen Wikipedia. Dabei werden die Kategorien jedoch unverändert in die Ontologie eingespeist

⁴⁷Ogleich auf S. 26 die Aussage getroffen wurde, dass die Darstellung von Kategoriebeziehungen als Baumstruktur irreführend ist, wurde hier eine Hierarchien implizierende Darstellung gewählt, damit die Problematik, sämtliche Beziehungstypen hierarchisch nennen zu wollen, deutlicher wird. Catgraph erlaubt auch eine Darstellung als „Wolke“ (vgl. [Catgraph 2009]).

⁴⁸Vgl. [DBpedia 2010].

⁴⁹Hier bezeichnet der Begriff Ontologie naturgemäß ein Begriffssystem im Allgemeinen und nicht, wie in Abschnitt 2.1, eine Dokumentationssprache.



Agnus Dei de.wikipedia super

und, wie in der Wikipedia, als über- beziehungsweise untergeordnete Kategorien bezeichnet. Eine Änderung oder Weiterentwicklung der Strukturen findet nicht statt.

In DBpedia übernommene Artikelnamen aus Wikipedia werden dort als „concepts“ bezeichnet, was hier mit „Begriffen“ übersetzt werden soll. Auf diesen Begriffen beruht der Bookmarking-Dienst Faviki⁵⁰, der es erlaubt, sie und damit ja Artikelnamen der Wikipedia zum Tagging von Bookmarks und also zur Verschlagwortung von Webseiten einzusetzen. Zumindest derzeit wird dabei innerhalb der Web-Benutzeroberfläche jedoch kein Gebrauch vom Kategoriensystem gemacht, obgleich über die DBpedia-Ontologie die benutzten Tags ja mit ihren Kategorien verbunden sind.

Zum anderen macht die Ontologie YAGO⁵¹ Gebrauch von den Kategorien der englischsprachigen Wikipedia. Dabei haben die Entwickler jedoch erkannt, dass die im Kategoriensystem ausgewiesenen Hierarchierelationen nicht immer solche sind – „the Wikipedia categories are indeed arranged in a hierarchy, but this hierarchy is barely useful for ontological purposes“ ([Suchanek 2007], S. 698) –, und beschränken ihre Arbeit auf diejenigen Kategorien, denen direkt Artikel zugeordnet sind, wobei die höheren Hierarchiestufen außen vor bleiben. Wird ein Artikelname als Begriff in die Ontologie aufgenommen, so werden die Namen der Kategorien, denen er direkt zugeordnet ist, linguistisch analysiert und zum einen auf Äquivalenzklassen aus der lexikalischen Datenbank WordNet – die wirklich hierarchisch aufgebaut ist –⁵² zurückgeführt und diesen untergeordnet, zum anderen auch zur Extraktion von Informationen genutzt, die dann auch in die Ontologie aufgenommen werden.⁵³ Wie bereits dargestellt, werden dabei jedoch nur die untersten Kategorien des Wikipedia-Systems untersucht, sämtliche Relationen innerhalb dieses Systems bleiben unberücksichtigt und kommen in YAGO nicht vor. Zudem setzt YAGO nach Untersuchungen des Autors dieser Arbeit den Fokus auf Individualbegriffe (Personen, Orte usw.) und enthält wenige Allgemeinbegriffe aus der Wikipedia. Die neu entstandenen YAGO-Kategorien (die untersten Wikipedia-Kategorien und ihre übergeordneten Äquivalenzklassen aus WordNet) erscheinen inzwischen auch im DBpedia-Projekt.

⁵⁰Vgl. [Faviki 2010]

⁵¹Vgl. [Suchanek 2007]. YAGO steht für „Yet Another Great Ontology“.

⁵²WordNet ist eine lexikalische Datenbank des Englischen, die Wörter ähnlicher Bedeutung zu sogenannten Äquivalenzklassen zusammenführt und semantische sowie lexikalische Beziehungen zwischen diesen ausweist. Vgl. die Website [WordNet 2010].

⁵³Dabei wird beispielsweise dem Kategoriennamen „1879 births“ die Information entnommen, dass eine Person, deren beschreibender Artikel dieser Kategorie zugeordnet ist, 1879 geboren wurde. Eine Beschreibung des Vorgehens gibt [Suchanek 2007] auf S. 701 – 703.

3.5 Fazit und Ausblick

In den vorigen Abschnitten wurden das Kategoriensystem der Wikipedia in seiner Entstehung beschrieben sowie sein derzeitiger Entwicklungsstand im deutschsprachigen Teil und sich daraus ergebende Problemstellungen. Dabei wurde zwischen durch Fachbereiche gepflegten Teilsystemen und solchen, die außerhalb von Fachbereichen liegen, unterschieden. Derzeitige Probleme sind insbesondere das Fehlen eines einheitlichen Regelwerks zur Kategorisierung im Gesamtsystem und sich daraus ergebende Inkonsistenzen, die vor allen Dingen außerhalb der Fachbereiche liegen, aber ebenso zwischen Fachbereichen, die nach unterschiedlichen Regeln kategorisieren.

Trotz dieser Probleme muss jedoch festgehalten werden, dass das Kategoriensystem der deutschsprachigen Wikipedia keineswegs als ein sinnloses Unterfangen oder als unbrauchbar zu bezeichnen ist, sondern das Potenzial zur Weiterentwicklung in Richtung einer höheren Leistungsstärke bietet. Aus der Tatsache, dass in den Fachbereichen in der Regel Fachleute des entsprechenden Gegenstandsbereichs arbeiten, ergibt sich die Vermutung, dass die meisten der vorhandenen Relationen in irgendeiner Hinsicht sinnvoll sind, auch wenn bei der Relationierung möglicherweise keine Informationswissenschaftler mitgewirkt haben und es derzeit nur einen Relationstypen gibt. Daneben verbietet es schon der zumindest in der deutschsprachigen Wikipedia erreichte Umfang des Kategoriensystems, es keiner weiteren informationswissenschaftlichen Betrachtung zu unterziehen und damit die an ihm Mitwirkenden Lügen zu strafen. Das Kategoriensystem der Wikipedia ist das bislang größte Projekt der gemeinschaftlichen Begriffsrelationierung, mithin also der gemeinschaftlichen Entwicklung eines Begriffssystems. Schon die schiere Zahl an Begriffen und Relationen fordert eine weitere Beschäftigung.

Benötigen manche Teile des Systems zwar noch grundlegende Überarbeitungen, so bieten zumindest die Teilsysteme einzelner Fachbereiche bereits jetzt die direkte Möglichkeit der Weiterentwicklung ihrer Strukturen unter Aufnahme des Bestehenden. Eine Weiterentwicklung bedeutet dabei in erster Linie eine differenzierte Typisierung der vorhandenen Relationen, die das Vorhandensein mehr als eines Relationstypen voraussetzt. Eine solche Weiterentwicklungsmöglichkeit eröffnet diese Arbeit praktisch in den folgenden Abschnitten.

4 Grundüberlegungen zur Entwicklung eines differenzierten Relationeninventars

Nachdem das Kategoriensystem der Wikipedia untersucht wurde, folgen in diesem Abschnitt nun Grundüberlegungen, die den Nutzen eines in Abschnitt 2.2 bereits angesprochenen differenzierten Relationeninventars sowie dessen Entwicklung betreffen. Diese Grundüberlegungen sind noch nicht auf das Kategoriensystem bezogen. Ein Beziehen der Überlegungen auf das Kategoriensystem erfolgt in Abschnitt 5.1.

4.1 Formulierung einer Struktur anhand eines Beispiels

Wie in Abschnitt 2.2 dieser Arbeit angedeutet wurde, ist es in vielen Retrievalkontexten sinnvoll und wünschenswert, zwischen Begriffen eines Begriffssystems nicht allein die drei beziehungsweise fünf klassischen Thesaurusrelationen zuzulassen, sondern diese zu verfeinern. Dieser Wunsch wird desto stärker, je umfangreicher ein Begriffssystem ist und je mehr Beziehungen von einem Begriff ausgehen. Hierfür sei ein Beispiel angegeben:

Man stelle sich den Begriff „Theater“ in einem fiktiven Begriffssystem, das nur einen Relationstypen kennt, vor, der durch diese eine völlige unspezifische Relation mit den folgenden Begriffen verbunden ist:

1. Politisches Theater
2. Alt-Wiener Volkstheater
3. Experimentelles Theater
4. Lustspiel
5. Bauerntheater
6. Theater am Küchengarten (in Hannover)
7. Bühne
8. Dramatiker
9. Schauspieler
10. Bühnentechnik
11. Schauspielschule
12. Theaterwissenschaft

Schaut man sich diese Begriffe unter dem Gesichtspunkt ihrer Verbindung zum Begriff „Theater“ an, ist zunächst festzuhalten, dass die Begriffe 1 bis 7 eine hierarchische Relation zu diesem haben – sind sie in der Thesaurusterminologie sämtlich Unterbegriffe von „Theater“ –, während die Relation der Begriffe 8 bis 12 zum Theaterbegriff assoziativ ist. Diese Unterscheidung formal festzuhalten, ist unter Verwendung der drei klassischen Relationstypen möglich.

Als nächstes kann die bekannte klassische Dreiteilung der Hierarchierelation durchgeführt werden: Bei den Begriffen 1 bis 5 liegen generische Unterbegriffsrelationen vor, bei Begriff 6 eine Instanzrelation, bei Begriff 7 eine partitive, insofern der Begriff des Theaters auch das Theatergebäude einschließt, von dem die Bühne naturgemäß ein Teil ist.

Bei noch genauerer Betrachtung zeigen sich jedoch große Unterschiede in den Aspekten, die die Über- und Unterordnung, insbesondere bei den generischen Relationen, beziehungsweise die nicht-hierarchische Verwandtschaft begründen. Bei Begriff 3 ist zu erkennen, dass sich die generische Unterordnung aus dem formalen Aufbau eines Theaterstücks ergibt: Experimentelles Theater ist Theater, das bestimmten formalen Aspekten folgt, die es von anderen Theaterformen abgrenzen. Bei den Begriffen 1 und 4 ergibt sich die Unterordnung aus dem Inhalt, also dem behandelten Stoff: Politisches Theater ist Theater, das bestimmten Stoff behandelt. Ein Lustspiel ist ebenso Theater, das bestimmten Stoff behandelt.⁵⁴ Das Bauerntheater hingegen wendet sich an eine bestimmte Zielgruppe und hat einen bestimmten Typus von Schauspielern. Beim Alt-Wiener Volkstheater treten lokale und historische Aspekte hinzu.

Ebenso große Unterschiede bestehen in den Aspekten, die die Assoziation bei den Begriffen 8 bis 12 begründen. Ein Dramatiker erschafft Theater, ein Schauspieler spielt Theater, die Bühnentechnik gestaltet Theater, eine Schauspielschule ist eine Institution, die Schauspieler ausbildet, und die Theaterwissenschaft betrachtet Theater von außen.

⁵⁴An dieser Stelle mag eingewendet werden, dass ein Lustspiel kein Theater, sondern ein Theaterstück sei. Diese Unterscheidung ist korrekt, führte im Kontext des Begriffssystems jedoch zu einer unnötigen Verkomplizierung: Zum einen ist bei einigen der anderen Begriffe eine Unterscheidung zwischen dem Theater an sich und einem Theaterstück nicht durchzuhalten. Ist ein Theaterstück, das den Regeln Experimentellen Theaters folgt, nun Experimentelles Theater oder vielmehr ein Experimentelles Theaterstück? Um die Unterscheidung zwischen Theater und Theaterstück konsistent durchführen zu können, müsste also neben „Experimentelles Theater“ ein weiterer Begriff, „Experimentelles Theaterstück“, eingeführt werden, der dann, ebenso wie „Lustspiel“, unterhalb eines ebenfalls neu eingeführten Begriffs „Theaterstück“ angesiedelt wird. Zum anderen soll, wenn das fiktive Begriffssystem für Verwendung durch Menschen, nicht durch Maschinen angelegt wird und also davon auszugehen ist, dass jeder Benutzer auch eine direkte Verbindung von „Theater“ und „Lustspiel“ erfassen und verstehen kann und sie ihm sogar lieber ist, da keine Zwischenhierarchie „Theaterstück“ eingeschaltet wird, hier einer pragmatischen Sichtweise gefolgt und diese direkte Verbindung auch so hergestellt werden. Festzuhalten ist also: *Der Begriff „Theater“ schließt das Theaterstück ein.* Wem diese Zusammenziehung logisch missfällt, der mag bei „Lustspiel“ jedoch an eine Aufführung eines Lustspiels (im Sinne eines Stücks) denken. Jene Aufführung ist ja auch ein Lustspiel, und sie ist Theater.

Im Folgenden werden die erkannten Beziehungstypen bezogen auf den Begriff „Theater“ aufgezählt:⁵⁵

1. Politisches Theater: Unterbegriff, generisch, Aspekt des Stoffs
2. Alt-Wiener Volkstheater: Unterbegriff, generisch, Aspekte des lokalen (geographischen) Bezugs, ebenso des zeitlichen (historischen) Bezugs
3. Experimentelles Theater: Unterbegriff, generisch, Aspekt des Aufbaus
4. Lustspiel: Unterbegriff, generisch, Aspekt des Stoffs
5. Bauerntheater: Unterbegriff, generisch, Aspekt der Beteiligten
6. Theater am Küchengarten: Unterbegriff, Instanz
7. Bühne: Unterbegriff, partitiv
8. Dramatiker: verwandter Begriff, erschaffende Person
9. Schauspieler: verwandter Begriff, darstellende Person
10. Bühnentechnik: verwandter Begriff, Aspekt der Gestaltung
11. Schauspielschule: verwandter Begriff, sonstige Beteiligung (hier: Lehre)
12. Theaterwissenschaft: verwandter Begriff, Betrachtung von außen

In dieser Aufstellung zeigt sich, dass zwölf verschiedene Typen von Relationen bestehen, die in ihrem detaillierten Bedeutungsgehalt durch die klassischen Thesaurusrelationen nicht beschrieben werden können. An dieser Stelle mag in einer pragmatischen Sichtweise eingewendet werden, dass bei einem Begriffssystem, das aus zwölf Begriffen besteht, eine feine Differenzierung der Relationen nicht nötig sei, da das System auch ohne eine solche Differenzierung überblickt werden könne. Dieser Einwand ist korrekt, er hat jedoch desto weniger Bestand, je größer das Begriffssystem und je größer die Zahl der Relationen werden, die von einem Begriff ausgehend ausgewiesen sind. Hat beispielsweise ein Begriff 50 Unterbegriffe, so ist es aus Benutzersicht wünschenswert, diese nach Aspekten gegliedert anzeigen und nutzen zu können, anstatt allenfalls eine alphabetische Liste angezeigt zu bekommen. Ein formuliertes differenziertes Relationeninventar dient also bei Begriffssystemen, die durch Menschen verwendet werden – an solche ist im Verlauf dieser Arbeit ausschließlich gedacht –, immer pragmatisch der Navigation und dem Retrieval. In diesen Bereichen muss es Vorteile bringen, die ein weniger differenziert typisiertes

⁵⁵ Aus Gründen der einfachen Formulierung und in Anlehnung an Thesauri wird hier der Ausdruck des verwandten Begriffs zur Bezeichnung der assoziativen Relation verwendet.

Relationeninventar oder, wie im oben dargestellten Extremfall, nur ein einziger Relationstyp, nicht bieten.

Werden Begriffsrelationen nun also differenziert typisiert, so ist jedoch zu beachten, dass eine zu kleinteilige Typisierung ebenfalls nicht wünschenswert ist. Im extremen Fall sind, wenn es viele Relationstypen gibt, die nur bei jeweils einem Begriffspaar erscheinen, sämtliche eben genannten Vorteile nicht mehr vorhanden: Wenn der Nutzer einen formal festgehaltenen Relationstypen betrachtet, der nur bei einem Begriffspaar vorkommt, so kann er ebenso das Begriffspaar direkt betrachten. In Begriffssystemen, die für eine Verarbeitung durch Maschinen gedacht sind, kann eine äußerst differenzierte Relationentypisierung von Nutzen sein, bei Begriffssystemen, die durch Menschen genutzt werden, ist jedoch immer ein mittlerer Weg zwischen genauer Typisierung und praktischem Nutzen zu wählen. Letzterer Effekt, der praktische Nutzen, kann jedoch auch umgekehrt vernichtet werden, wenn Relationstypen zu sehr zusammengefasst werden und damit von der eigentlichen individuellen Beziehung im Begriffspaar zu stark abstrahiert wird. Dann besteht die Gefahr, dass die Relationen nur noch durch sehr abstrakte Begriffe beschrieben werden, die wiederum für viele Nutzer nicht mehr verständlich wären, da sie von dem eigentlichen Bedeutungsgehalt einer individuellen Relation zu weit wegführten.

In der Relationentypisierung existiert zusammengefasst also ein Spannungsfeld zwischen der korrekten Abbildung des semantischen Gehalts einer Begriffsrelation, dem Abstraktionsgrad und der Verständlichkeit der schließlich formulierten Bezeichnung der Relation. Wie in allen Aspekten der Gestaltung von Begriffssystemen kann es auch hier keine allgemeingültige Antwort geben. Das Relationeninventar eines Begriffssystems muss wie dessen gesamte Gestaltung immer von der intendierten Nutzung abhängig gemacht werden.

Wie bereits in Abschnitt 2.2 dargestellt, kann die Verständlichkeit fein typisierter Relationen durch die Nutzer in hohem Maße gesteigert werden, wenn die neuen Relationen als Verfeinerungen der klassischen begriffen werden. Das bedeutet, dass das Relationeninventar eines Begriffssystems seinerseits hierarchisch geordnet wird. Jede neue Relation ist eine Verfeinerung einer der drei klassischen Relationen. Dieses Vorgehen ist möglich, da, wie bereits erwähnt, die drei klassischen Relationstypen bereits jede mögliche Relation enthalten, jedoch naturgemäß in sehr stark abstrahierter Form: Jede Relation zwischen zwei Begriffen ist entweder eine hierarchische, eine Äquivalenzrelation oder nichts von beidem und damit assoziativ.

Durch eine hierarchische Ordnung fällt es dem Nutzer leichter, das Relationeninventar zu durchdringen und zu verstehen, da er die neuen Relationen eben als Verfeinerungen der klassischen begreift und nicht etwa denken kann, dass sie neben den alten stünden. Dabei kann es auch mehrere Stufen der Verfeinerung geben, wodurch mehrstufige Hierarchieleitern innerhalb des Inventars entstehen. Die oben für die beispielhaften 12 Begriffe formulierten Relationstypen können in folgende

Hierarchie gebracht werden:

1. hierarchische Relation
 - a) generische Relation
 - i. nach Aspekt der Form
 - A. nach Aspekt des Aufbaus
 - B. nach Aspekt des behandelten Stoffs
 - ii. nach Aspekt der Beteiligten
 - iii. nach Aspekt des lokalen Bezugs
 - iv. nach Aspekt des zeitlichen Bezugs
 - b) partitive Relation
 - c) Instanzrelation
2. Äquivalenzrelation
3. assoziative Relation
 - a) Beteiligung
 - i. erschaffende Person
 - ii. gestalterisches Konzept
 - iii. darstellende Person
 - iv. betrachtendes Konzept

So ist ein erster Entwurf eines Relationeninventars für den Gegenstandsbereich Theater entstanden. Einige Anmerkungen hierzu:

- Bei der generischen Relation mag die Zusammenfassung der Aspekte Aufbau und Stoff zu einem übergeordneten Aspekt Form zunächst künstlich beziehungsweise unverständlich erscheinen. Diese Einordnung ist darauf zurückzuführen, dass sich die Aspekte Aufbau und Stoff auf ein Theaterstück beziehen, also auf die grundlegende Form des aufgeführten Theaters, ja das, was das Theater an sich ausmacht. Der Aspekt der Beteiligten ist davon unabhängig, denn eine Gruppe von Beteiligten an sich kann naturgemäß Theater jeglicher Form betreiben oder betrachten. Das Gleiche gilt für lokale und zeitliche Bezüge.
- Die assoziative Relation enthält als einzigen direkt untergeordneten Gliederungspunkt bisher die, noch schwammig formulierte, „Beteiligung“. Dieser Punkt ist bereits jetzt eingefügt worden, da davon auszugehen ist, dass auch eine Vielzahl anderer assoziativer Beziehungen existiert, die nicht auf eine Beteiligung am Theater zurückzuführen sind.

- Die Relation der „Beteiligung“ ihrerseits ist nicht etwa nach Personen und anderem (etwa Gruppen oder Institutionen) gegliedert, sondern hat zunächst vier Verfeinerungen nach Erschaffung, Gestaltung, Darstellung und Betrachtung. Diese Gliederung erscheint sinnvoll, da sie vier grundlegende Typen der Beteiligung am Theater beschreibt: Die Erschaffung (ein kreativer Prozess, der etwa in Autoren von Theaterstücken abläuft), die Gestaltung (die Tätigkeiten etwa von Bühnentechnikern und Regisseuren), die Darstellung (die Tätigkeit etwa von Schauspielern) sowie die Betrachtung (die Tätigkeiten von Zuschauern, aber auch von Kritikern oder der Theaterwissenschaft). Unterhalb dieser vier Relationstypen ist eine Gliederung in Personen und Konzepte angedeutet. Der Begriff des Konzepts ist naturgemäß schwammig und vieldeutig. Er wurde eingeführt zur Unterscheidung zwischen einer beteiligten Person, etwa einem Schauspieler, und etwas beteiligtem Abstraktem, wie der Bühnentechnik (als Gesamtheit der technischen Vorgänge, nicht als Bezeichnung für eine Gruppe von Bühnentechnikern) und der Theaterwissenschaft.
- Die eingangs als „sonstige Beteiligung“ formulierte Relation zwischen der Schauspielschule und dem Theater ist nicht explizit formuliert. Dies ergibt sich aus dem Anwendungsprinzip begriffsorientierter Dokumentationssprachen der Vergabe des genauesten Begriffs einer Hierarchieleiter. Da eben kein genauere Relationstyp vorhanden ist, würde die Beziehung zwischen Schauspielhaus und Theater schlicht als „Beteiligung“ angegeben werden.

Diese praktischen Überlegungen sollen die Vorgänge in der Formulierung hierarchisch gegliederter Relationeninventare anhand eines Beispiels verdeutlichen.

Es fällt jedoch auf, dass das so beispielhaft entwickelte Inventar in hohem Maße auf den Gegenstandsbereich des Theaters zugeschnitten ist. Dies führt eine weitere Problemstellung im oben genannten Spannungsfeld vor Augen: Wird ein Relationeninventar nur für einen umgrenzten Gegenstandsbereich entwickelt, so können die Relationstypen in ihrer Formulierung nah am konkreten semantischen Gehalt der Beziehungen bleiben. Wird ein einziges Inventar für ein themenübergreifendes Begriffssystem formuliert, so müssen notwendigerweise abstraktere Formulierungen getroffen werden, die aufgrund ihres Abstraktionsgrades bei der konkreten Anwendung in einem umgrenzten Gegenstandsbereich möglicherweise unverständlich werden.

In der bereits in Abschnitt 2.2 vorgestellten Arbeit von Boteram wird versucht, diese Spannung aufzulösen, indem in einem mehrstufig gegliederten Relationeninventar die Relationen auf den oberen Stufen für alle Gegenstandsbereiche gelten und damit abstrakt sind, während, wenn nötig, unterhalb einer Stufe weitere, feine und themenspezifische Relationstypen mit ausschließlich themenspezifischer Gültigkeit angekoppelt werden, sofern sich die allgemeineren als zu abstrakt für einen Gegenstandsbereich erweisen. So führt Boteram eine allgemeingültige Relation „Un-

terbegriff, generisch, nach Aspekt der Form“ ein, die im Gegenstandsbereich Theater weiter nach den Aspekten „Genre“ und „Tradition“ verfeinert werden soll.⁵⁶ Boteram geht davon aus, dass die Relationen auf den drei oberen Stufen allgemeingültig sein sollen, während darunter gegenstandsspezifische Verfeinerungen erfolgen können.

Eine andere Möglichkeit der Lösung des Problems mangelnder Allgemeingültigkeit von Relationen besteht eben darin, kein themenübergreifendes Relationeninventar zu formulieren, sondern vollkommen spezifische Inventare für bestimmte Gegenstandsbereiche zu entwickeln, bei denen allenfalls die oberste Stufe mit den drei klassischen Relationstypen sowie im Fall der hierarchischen Relation gegebenenfalls die zweite Stufe und also die generische, die partitive und die Instanzrelation übereinstimmen.

Diese Arbeit möchte keine endgültige Aussage darüber treffen, welche der beiden Vorgehensweisen die sinnvollere ist. Es wird jedoch versucht, zunächst ein Inventar für einen bestimmten Gegenstandsbereich, das Theater, zu entwickeln und dann durch eine Anwendung auf einen anderen Gegenstandsbereich, die Jagd, Erkenntnisse darüber zu gewinnen, welche der also zunächst theaterspezifisch entwickelten Relationen allgemein anwendbar sind.

4.2 Nutzen eines differenzierten Relationeninventars

Welchen Nutzen ein wie im vorigen Abschnitt skizziert gestaltetes Relationeninventar aus der Sicht der Anwender bietet, stellt [Boteram 2008] auf S. 35–38 dar. Die Nutzungsmöglichkeiten sollen im Folgenden knapp wiedergegeben werden.

Ein Nutzen besteht in den erweiterten Möglichkeiten in der Orientierung im Vokabular, also in der Navigation. Wie bereits erwähnt, können ohne eine Relationendifferenzierung die Unterbegriffe eines Begriffs allenfalls alphabetisch sortiert wiedergegeben werden. Das Gleiche gilt für die verwandten Begriffe. Stehen verfeinerte Relationstypen zur Verfügung, ermöglicht dies beispielsweise eine nach den jeweiligen Aspekten der Unterordnung gegliederte Anzeige der Unterbegriffe. Es kann also anhand einzelner Aspekte navigiert, das Vokabular also unter bestimmten Gesichtspunkten durchforstet werden. Diese Möglichkeit wird desto wichtiger, je mehr Relationen von einem Begriff ausgehen.

Sofern es sich bei dem Begriffssystem um eine Dokumentationssprache handelt, also um ein Instrument der Dokumenterschließung und -suche, ergeben sich durch differenzierte Relationstypen spezifische Möglichkeiten der Modifikation und Erweiterung von Suchanfragen. So wird es möglich, in eine Suche nur diejenigen Unterbegriffe eines Begriffs einzubeziehen, bei denen die Unterordnung einem bestimmten Aspekt folgt. Ebenso kann eine zu große Treffermenge nach bestimmten Aspekten verkleinert, eine zu kleine nach bestimmten Aspekten vergrößert werden.

⁵⁶Vgl. [Boteram 2008], S. 66–73.

4.3 Entwicklung eines Relationeninventars anhand bestehenden Vokabulars

In Abschnitt 4.1 wurde beispielhaft ein neues Relationeninventar anhand eines Systems von zwölf Begriffen geschaffen. Es ist wichtig, sich an dieser Stelle bewusst zu machen, dass das neue Inventar also anhand eines bereits bestehenden Vokabulars entwickelt wurde, mithin also auf dieses Vokabular zugeschnitten ist.

Eine andere Möglichkeit des Vorgehens ist die Entwicklung eines Relationeninventars – sei es themenübergreifend oder auf einen Gegenstandsbereich zugeschnitten –, unabhängig von einem bestimmten Vokabular und ohne ein Vokabular vor Augen. So ließe sich beispielsweise ein Relationeninventar zum Gegenstandsbereich Theater entwickeln, das in allen Vokabularen zu diesem Bereich Anwendung finden könnte. Ebenso ließe sich ein themenübergreifendes Inventar entwickeln, das nun also überall Anwendung finden könnte.

Eine solche Vorgehensweise hat den Vorteil, dass die Anwender des Vokabulars nur den Umgang mit einem Relationeninventar lernen müssten. Gäbe es ein vokabularübergreifendes Inventar zum Thema Theater, so müsste ein Theaterwissenschaftler dessen Struktur nur einmal erlernen und könnte in jeglichen Vokabularen des Gegenstandsbereichs navigieren. Gäbe es sogar ein themenübergreifendes und allgemeingültiges Inventar, so müsste jeder Nutzer nur einmal dessen Struktur erlernen, so wie es derzeit für die Nutzung der meisten Vokabulare nötig ist, die drei klassischen Relationstypen zu kennen.

Vokabularübergreifende Relationeninventare haben jedoch den Nachteil, dass in ihnen die Schwerpunkte einzelner Vokabulare nicht in dem Maße berücksichtigt werden können wie in vokabularspezifischen Inventaren. Man stelle sich ein Vokabular zum Thema Theater vor, das differenzierte Begriffe und Begriffsrelationen zu Aspekten der Theatergeschichte enthält, mithin also einen historischen Fokus hat. In einem solchen Vokabular kann es nötig werden, die oben „lokaler Bezug“ und „zeitlicher Bezug“ benannten Relationstypen weiter zu differenzieren. Dies wäre nicht möglich, wenn es ein feststehendes Inventar zum Thema Theater gäbe, dessen Status und Verbreitung keine Änderungen zuließen.

Um eine größtmögliche Spezifität der im Inventar enthaltenen Relationstypen zu gewährleisten, ist also der Zuschnitt auf ein bestimmtes Vokabular vorzuziehen. Dies bedeutet mithin also, dass das Vokabular oder zumindest eine unrelationierte Begriffssammlung bereits, zumindest in seinen beziehungsweise ihren Grundzügen, bestehen müssen. Die Entwicklung des Inventars erfolgt also, indem, wenn noch nötig, zunächst undifferenzierte Relationen ausgewiesen und diese daraufhin in ihrem semantischen Gehalt möglichst genau untersucht werden. So entsteht zunächst eine Vielzahl von Relationstypen, die sehr dicht an der konkreten Bedeutung einer Bezie-

hung sind. Da jedoch, wie im Abschnitt 4.1 dargelegt, aus Gründen der Praktikabilität ein bestimmter Abstraktionsgrad in den Relationstypen vonnöten ist, müssen diese sehr spezifischen Relationstypen daraufhin zu einem gut handhabbaren Inventar zusammengefasst werden. Ein solches Inventar wird jedoch immer stärker an den konkreten Bedeutungen der vorhandenen Relationen bleiben als eines, das nicht aus dem bestehenden Vokabular entwickelt wurde.

Ein vokabularspezifisches Inventar ermöglicht zudem eine Anpassung an den intendierten Nutzerkreis des Vokabulars. Ein Begriffssystem, das für eine Verwendung durch Fachwissenschaftler vorgesehen ist, kann und muss differenziertere Relationstypen enthalten als eines, das für eine nicht-berufliche Verwendung durch einen unumrissenen Nutzerkreis gedacht ist. Wird das Vokabular etwa von Personen, die keine fachlichen Kenner des dargestellten Gebiets sind, in ihrer Freizeit genutzt, so können zu differenzierte Relationstypen gar zu Überforderung und Unsicherheit führen. In noch höherem Maße gilt dies, wenn das Inventar durch einen unumrissenen Kreis von Personen selbst zur Relationierung angewandt werden soll, wie es im Verlauf dieser Arbeit vorgeschlagen wird.

Zusammenfassend ist ein vokabularspezifisches Inventar daher einem vokabularübergreifenden vorzuziehen. So soll in dieser Arbeit auch ein vokabularspezifisches entwickelt werden, das auf das Kategoriensystem der Wikipedia zugeschnitten ist.

5 Entwicklung eines Relationeninventars zur gemeinschaftlichen Anwendung im Kategoriensystem der Wikipedia

Im folgenden Abschnitt wird nach einigen Grundüberlegungen, die nun im Speziellen das Kategoriensystem der Wikipedia betreffen, die im Rahmen dieser Arbeit erfolgte Entwicklung eines Relationeninventars zur gemeinschaftlichen Anwendung im Kategoriensystem der Wikipedia beschrieben. Hierzu werden zunächst zwei Inventare für zwei Gegenstandsbereiche entwickelt. Diese beiden themenspezifischen Inventare sind in den Anhängen A beziehungsweise B tabellarisch dargestellt. Die mit ihnen relationierten Vokabulare sind aufgrund ihrer Größe nicht abgedruckt, sondern auf der dieser Arbeit beigelegten CD-ROM in den Verzeichnissen *theater/* beziehungsweise *jagd/* enthalten. Aus den zwei Inventaren wird daraufhin ein themenübergreifendes Inventar entwickelt, das im Text dieses Abschnitts vollständig dargestellt und beschrieben wird.

5.1 Wikipedia-spezifische Grundüberlegungen

Im vorigen Abschnitt wurden Grundüberlegungen betreffend Relationeninventare im Allgemeinen getätigt. Auf den Ergebnissen jener Überlegungen baut das zu entwickelnde Inventar für das Kategoriensystem der Wikipedia auf. Zusammenfassend bedeutet dies, dass das Inventar hierarchisch geordnet sein und einen vokabularspezifischen Charakter haben soll: *Es ist für die Verwendung im Kategoriensystem der Wikipedia gedacht.* Anwendungen darüber hinaus sind, zumindest zu diesem Zeitpunkt, nicht vorgesehen.

Aus dem Charakter des Kategoriensystems ergeben sich zwei weitere Problemstellungen, zu denen vor Beginn der Entwicklung Überlegungen vonnöten sind:

1. *Das Kategoriensystem wird kollaborativ aufgebaut.* Das bedeutet, dass das zu entwickelnde Inventar kollaborativ durch eine nicht näher umrissene Personengruppe angewandt werden soll, nämlich durch jeden, der an der Entwicklung der Wikipedia beteiligt ist, mithin also potenziell jeden Internetnutzer. Dies stellt besondere Anforderungen an die Allgemeinverständlichkeit der neuen Relationstypen. Es ist im skizzierten Spannungsfeld in der Inventarentwicklung ein Weg der einfachen Verständlichkeit und eines niedrigen Komplexitätsgrades zu wählen.⁵⁷

⁵⁷An dieser Stelle sei darauf hingewiesen, dass der Erfolg der Wikipedia nicht zuletzt auf der geringen Komplexität und der allgemeinen und intuitiven Verständlichkeit des Wiki-Prinzips beruht. Frost schreibt hierzu: „Es hat Züge von Ironie, dass Wissenschaftler in internationalen Projekten komplexe Wissensmanagementsysteme entwickeln, sich aber parallel dazu eine minimalistische, von einer Privatperson erfunden[e] Software durchsetzt.“ ([Frost 2006], S. 120.) Diese Aussage ist in ihrer Formulierung nicht haltbar, da sie suggeriert, dass komplexere Wissensmanagementsysteme generell unnötig wären – vielmehr gibt es naturgemäß durchaus Anwendungszwecke,

Dies bedeutet zunächst, eine nicht zu hohe Zahl an Relationstypen zu schaffen. Gibt es zu viele Relationstypen, so ist, auch wenn diese hierarchisch angeordnet sind, die Allgemeinverständlichkeit der einzelnen Typen sowie der Bedeutungsunterschiede zwischen ihnen nicht mehr gewährleistet. Michel formuliert in seiner Relationentaxonomie [Michel 1997] mehr als 100 verschiedene Relationstypen. Eine so hohe Zahl kann von nicht informationswissenschaftlich ausgebildeten Fachleuten nicht mehr überblickt werden und eignet sich daher nicht für eine kollaborative Anwendung. Im vorliegenden Anwendungsszenario ist also die Zusammenfassung einzelner bedeutungsähnlicher Relationstypen zu einem einzelnen vonnöten, was naturgemäß mit einem höheren Abstraktionsgrad, jedoch auch mit einem höheren Nutzen in der kollaborativen Anwendung einhergeht.

Ein zweites Problem innerhalb dieser Problemstellung ergibt sich jedoch genau aus dem zu wählenden Abstraktionsgrad: Es ist darauf zu achten, dass die notwendigerweise in einem gewissen Maß abstrakten Relationstypen nicht durch zu abstrakte Bezeichnungen ausgedrückt werden, da bei solchen ebenfalls keine allgemeine Verständlichkeit gewährleistet ist. Dieses Problem wird insbesondere dann deutlich, wenn das zu entwickelnde Inventar themenübergreifend sein, mithin also im konkreten Anwendungsfall für das gesamte Kategoriensystem der Wikipedia gelten soll. Dies führt zur im Folgenden erläuterten zweiten Problemstellung.

2. *Das Kategoriensystem ist themenübergreifend.* Es ist zu überlegen, ob also ein einziges, zwar vokabularspezifisches, aber themenübergreifendes Inventar gestaltet werden soll, das dann Anwendung im gesamten Kategoriensystem findet – was notwendigerweise mit einem hohen Abstraktionsgrad der Relationstypen einhergeht –, oder ob für einzelne Themenbereiche, etwa die durch die Fachbereiche abgegrenzten, jeweils spezifische Inventare entwickelt werden sollen, wobei die Relationstypen naturgemäß näher an den konkreten Bedeutungsgehalten der Relationen bleiben.⁵⁸ Auf diese Frage möchte die vorliegende Arbeit keine endgültige Antwort geben. Vielmehr soll, wie bereits angedeutet, versucht werden, sich einer Antwort zu nähern, indem zunächst ein themenspezifisches Vokabular für einen Gegenstandsbereich entwickelt wird. Dieses wird daraufhin auf einen anderen Gegenstandsbereich bezogen, der keinerlei

die mit dem Wiki-Prinzip nicht abgedeckt werden können –, weist jedoch auf die folgende Tatsache hin: Eine Sache, die über eine fachbezogene Personengruppe hinaus erfolgreich sein soll, muss intuitiv verständlich sein. Diese Verständlichkeit darf, wo nötig, durchaus pragmatisch zulasten einer übermäßig hohen Detailgenauigkeit gehen. Dies gilt umso mehr für das Relationeninventar.

⁵⁸Im WikiProjekt Kategorien wird beschrieben, dass sich die Kategorisierungen und damit die Relationierungen innerhalb eines Fachbereichs an den Bedürfnissen des jeweiligen Fachbereichs orientieren sollen und hierfür keine fachbereichsübergreifenden Regeln aufgestellt werden können (vgl. das Zitat auf S. 27). Dies könnte möglicherweise auch für Relationeninventare gelten.

Überschneidungen mit ersterem aufweist. Aus dieser Vorgehensweise werden Erkenntnisse hinsichtlich dieser Problemstellung erhofft.

Außerdem ist vorher festzulegen, als was das Kategoriensystem beziehungsweise wie dessen Umfang begriffen wird. Wie in Abschnitt 3 dargestellt, gibt es hierzu zwei Sichtweisen: Einerseits kann das Kategoriensystem als ein Begriffssystem angesehen werden, das nur die Kategoriennamen selbst als Begriffe enthält, und die Artikel somit durch die Kategorienbegriffe verschlagwortet würden, andererseits kann es um die Artikelnamen erweitert werden, die also Teil des Begriffssystems würden und jeweils an den Enden der Netzstrukturen stünden. Erstere Sichtweise bedeutet, dass nur die Relationen zwischen zwei Kategoriennamen typisiert würden, letztere, dass auch die Relationen zwischen Kategorien- und Artikelnamen typisiert würden.

Sofern Letzteres getan wird, sind die Vorteile hiervon zu benennen, da ein solches Vorgehen naturgemäß einen stark erhöhten Aufwand bedeutet. Um jene Vorteile zu begreifen, muss zunächst erkannt werden, dass im Gegensatz zu Dokumentations-sprachen, die der Erschließung von Literatur dienen, hier auch das zu Erschließende oftmals den Charakter von Begriffen hat: Viele Artikel beschreiben Objekte im Sinne von Wersig, also geistige Vorstellungen oder Gegenstände.⁵⁹ Das heißt, dass ihre Namen Begriffe darstellen, die in Deskriptoren überführt werden können.⁶⁰ Der Vorteil eines solchen Vorgehens liegt in einer verbesserten Navigation innerhalb der Wikipedia: Es kann so auch innerhalb einer Kategorie nach Aspekten navigiert werden. Viele Kategorien enthalten derzeit eine hohe Zahl an Artikeln. Können diese nach dem Aspekt ihrer Einordnung in die betreffende Kategorie sortiert werden, so bringt dies einen großen Nutzen in der Navigation. Deswegen soll hier so verfahren werden. Es werden im Verlauf dieser Arbeit also sowohl die Kategorien- als auch fast alle Artikelnamen als Begriffe verstanden und die Relationen zwischen ihnen typisiert. Nicht eingeschlossen werden nur diejenigen Artikelnamen, die keinen Begriff bezeichnen, etwa die „Liste von Operetten“ oder das „Ballett-Glossar“.

Nicht eingeschlossen werden in dieser Arbeit ebenfalls die bereits beschriebenen Artikelweiterleitungen, also Synonyme zu Artikelnamen. Da zu Kategoriennamen derzeit keine Synonyme gespeichert werden können und eine Einbeziehung der Artikelsynonyme also lediglich bedeuten würde, dass, betrachtet man das gesamte Begriffssystem, lediglich an den Enden der Netzstrukturen einige Äquivalenzbeziehungen vorhanden wären, schien es dem Autor dieser Arbeit konsequenter, diese Beziehungen völlig auszuklammern. In der vorliegenden Arbeit werden also keinerlei Äquivalenzbeziehungen betrachtet. Sollte es in Zukunft möglich werden, auch Weiterleitungen auf Kategorien zu speichern, so sollte das gestaltete Begriffssystem um

⁵⁹Vgl. [Wersig 1978], S. 13f.

⁶⁰Dies gilt naturgemäß nicht für Artikel, die kein Objekt beschreiben, sondern etwa „Liste von Operetten“ oder „Ballett-Glossar“ heißen. Diese Artikelnamen sind von der Einbettung in das Begriffssystem natürlich auszuschließen.

Äquivalenzbeziehungen an jeglicher Stelle, also bei Kategorien *und* Artikeln, erweitert werden.

5.2 Beschreibung sämtlicher Arbeitsschritte

Wie bereits dargestellt, erfolgt die Entwicklung des Relationeninventars in dieser Arbeit anhand eines bereits bestehenden Vokabulars. Dies ist das Vokabular zum Gegenstandsbereich Theater im Kategoriensystem der deutschsprachigen Wikipedia. Dieser Gegenstandsbereich wurde zum einen gewählt, da er bereits im Projekt MACS bearbeitet wurde⁶¹, und zum anderen, da das in [Boteram 2008] entwickelte Relationeninventar ebenfalls für diesen Bereich, jedoch in einem anderen Vokabular, der Schlagwormdatei, ist. Eine Bearbeitung des gleichen Gegenstandsbereichs in einem anderen Begriffssystem erlaubt später also einerseits Vergleiche betreffend die Repräsentation des Bereichs in unterschiedlichen Begriffssystemen und andererseits einen Vergleich des von Boteram entwickelten Inventars mit dem hier entwickelten. Solche Vergleiche sind jedoch nicht mehr Teil dieser Arbeit.

Das für das Theater entwickelte Relationeninventar wird zunächst sehr spezifischer Natur und lediglich für diesen Gegenstandsbereich, gegebenenfalls auch für verwandte, wie etwa den Film, geeignet sein. Nach seiner Entwicklung wird versucht, es auf einen völlig anderen Gegenstandsbereich, nämlich den der Jagd zu beziehen. Dieser Bereich wiederum wurde gewählt, da im Rahmen des Projektes CrissCross bereits seine Strukturen in der Schlagwortnormdatei sowie in der Dewey-Dezimalklassifikation (DDC) untersucht wurden und in [Hubrich 2009] kurz dargestellt werden, und somit auch diesbezügliche Vergleiche mit dem Wikipedia-Kategoriensystem später möglich sind. Anhand der Anwendung des theaterspezifischen Inventars auf das Thema Jagd soll herausgearbeitet werden, welche Teile des Theaterinventars in sich überschneidenden Bedeutungsinhalten auch im Bereich Jagd erscheinen. Indem nun also von den für den theaterspezifischen Relationstypen so weit abstrahiert wird, dass sie auch für den Bereich Jagd genutzt werden können, lassen sich erste Schlüsse in der Frage ziehen, ob es möglich ist, ein themenübergreifendes Inventar zu erschaffen, das jedoch nicht zu abstrakt formuliert ist, um nicht, wie bereits dargestellt, zumindest in kollaborativer Relationierungsarbeit gar nicht oder nur noch schwer anwendbar zu sein.

Als Grundlage für beide Untersuchungen, sowohl im Theater- als auch im Jagdbereich, dient die deutschsprachige Wikipedia nach dem Stand vom 3. Juni 2010. An diesem Tag wurde der zum Zeitpunkt der Durchführung dieser Arbeit neueste Dump der deutschsprachigen Wikipedia zur Verfügung gestellt. Ein Dump ist ein Abzug sämtlicher in Wikipedia enthaltener Daten. Dumps können, getrennt nach techni-

⁶¹Das Projekt MACS (Multilingual Access to Subjects) hat zum Ziel, Verbindungen zwischen drei bibliothekarischen Vokabularen unterschiedlicher Sprachen herzustellen (vgl. [MACS 2010]).

schen beziehungsweise formalen Bereichen der Wikipedia, frei heruntergeladen und verwendet werden. Das genaue technische Vorgehen wird in Abschnitt 5.3.3 erläutert.

Zu beachten ist, dass die Entwicklung des Inventars ausschließlich anhand der bereits vorhandenen Begriffsrelationen, also den vorhandenen Einordnungen von Kategorien und Artikeln in Kategorien erfolgt, also nur die vorhandenen Relationen untersucht werden. Es werden keine Relationen hinzugefügt, ebenso wird keine Relation entfernt, auch wenn sie sinnlos oder falsch ist.⁶²

5.3 Exemplarische Relationierung des Gegenstandsbereichs Theater der deutschsprachigen Wikipedia

Als zum Gegenstandsbereich Theater gehörig werden sämtliche Kategorien und Artikel betrachtet, die unterhalb der Kategorie „Theater“ liegen, unabhängig davon, wie viele Stufen sie von ihr entfernt sind. Dabei werden lediglich diejenigen ihrer Beziehungen betrachtet, die auf die Kategorie „Theater“ als oberste Kategorie zulaufen. Daneben gibt es aufgrund der themenübergreifenden Netzstruktur des Kategoriensystems naturgemäß auch Beziehungen derselben Kategorien und Artikel, die von der Kategorie „Theater“ wegführen und auf andere Teile des Systems zulaufen. Jene Beziehungen werden hier jedoch nicht betrachtet.

5.3.1 Derzeitige Struktur

Zum Zeitpunkt der Betrachtung liegen unterhalb der Kategorie „Theater“ 194 weitere Kategorien, davon sind ihr 24 direkt untergeordnet. In sämtlichen der 194 Kategorien sowie der Kategorie „Theater“ selbst sind zusammen 12.769 Artikel eingeordnet. Da einige Kategorien und Artikel mehreren Kategorien unterhalb der Theaterkategorie zugeordnet sind, gibt es insgesamt 15.809 Relationen. Gemäß der Natur eines enzyklopädischen Lexikons und insbesondere der Wikipedia ist in ihr jedoch eine Vielzahl von Artikeln über Individualbegriffe – hierbei konkrete Personen, Theaterensembles, Theatergebäude und so weiter – enthalten, sodass 14.680 der genannten Relationen Beziehungen zu Instanzen sind.⁶³ Es bleiben also 1.129 Beziehungen, die keine solchen sind, und daher im Kern einer Relationenbetrachtung unter semantischen Gesichtspunkten stehen sollen.

⁶²In der Praxis war nur ein sehr geringer Teil der Relationen offensichtlich falsch. Es gibt in den zwei untersuchten Kategorienteilsystemen also nur eine sehr geringe Zahl an Fehlzuordnungen.

⁶³Dieser Wert wurde erst im Rahmen der Relationierungsarbeit ermittelt, soll an dieser Stelle jedoch schon genannt werden, um dem Leser einen Überblick über die Struktur des Gegenstandsbereichs zu ermöglichen. Bei den genannten Relationen handelt es sich jedoch zum einen um hierarchische Instanzrelationen, zum anderen um assoziative Beziehungen, bei denen ein Partner im Begriffspaar eine Instanz bezeichnet.

```

Theater NT Ballett
Theater NT Drama
Theater NT Liste_(Theater)
Theater NT Musiktheater
Theater NT Puppentheater
Theater NT Schauspielschule
Theater NT Tanztheater
Theater NT Theater_nach_Genre
Theater NT Theater_nach_Ort
Theater NT Theater_nach_Staat
Theater NT Theaterbegriff
Theater NT Theaterbetrieb
Theater NT Theaterensemble
Theater NT Theaterfestival
Theater NT Theaterfotografie
Theater NT Theatergebäude
Theater NT Theaterkritik
Theater NT Theatermusik
Theater NT Theatername
Theater NT Theaterorganisation
Theater NT Theaterpreis
Theater NT Theaterschaffender
Theater NT Theaterverlag
Theater NT Theaterwissenschaft

```

Abbildung 4: Direkt untergeordnete Kategorien der Kategorie „Theater“ im bisherigen Stand

Die 24 Kategorien, die der Kategorie „Theater“ direkt untergeordnet sind, sind in Abbildung 4 dargestellt.⁶⁴

Es zeigt sich bereits hier die Existenz einer Vielzahl von Beziehungstypen. Während Ballett, Drama, Musiktheater, Puppentheater und Tanztheater beispielsweise als Sparten des Theaters zu bezeichnen sind, Theaterbetriebe, Theaterensembles und Theaterschaffende am Theater beteiligt sind, bezeichnen die Kategorien „Theater nach Genre“, „Theater nach Ort“ sowie „Theater nach Staat“ keine Begriffe, sondern dienen lediglich der Ordnung von Instanzen nach den drei genannten Aspekten. Bei den anderen Kategorien ergeben sich Unterordnungen oder assoziative Relationen verschiedenster Typen zum Begriff Theater.

Unklar ist zunächst im Besonderen die Bedeutung der Kategorie „Theaterbegriff“, deren Name ja keinen Begriff im Sinne von [Wersig 1978] bezeichnet. Die Struktur unterhalb dieser Kategorie wird in Abbildung 5 gezeigt, wobei hier neben den Kate-

⁶⁴Die Darstellung wurde mit dem in Abschnitt 5.3.3 beschriebenen Skript `browse-down` erzeugt. Die Zeilen entsprechen dem Format „*Begriff_1* NT *Begriff_2*“, was ausdrücken soll, dass zwischen *Begriff_1* und *Begriff_2* eine Beziehung vom Typ NT („narrow term“ beziehungsweise Unterbegriff) besteht. Die Bezeichnung der Beziehung erfolgt, wie im genannten Abschnitt erläutert wird, hier analog zu den in der Wikipedia verwendeten Begrifflichkeiten der Über- und Unterordnung von Kategorien, obgleich eine Über- und Unterordnung ja nicht bei allen Beziehungen vorliegt.

gorien auch die direkt untergeordneten Artikel einbezogen sind, die im Gegensatz zu jenen mit dem Präfix „a:“ gekennzeichnet sind. Es zeigt sich, dass die Kategorie eine Sammelkategorie für jegliche Begriffe aus dem Umfeld des Theaters ist. Auffallend sind hier vor allen Dingen einige der direkt untergeordneten Kategorien, etwa „Theatergenre“ und „Theaterberuf“, die ebenfalls selbst keine Begriffe bezeichnen, sondern lediglich eingeführt wurden, um ihrerseits Begriffe zu sammeln, die in einer bestimmten Beziehung zum Begriff Theater stehen, nämlich einerseits Genres, die nach den Überlegungen in Abschnitt 4.1 ja generische Unterbegriffe unter anderem nach den Aspekten Aufbau oder Stoff sein können, und andererseits Berufsbezeichnungen im Theaterzusammenhang, die ja verwandte Begriffe nach dem Aspekt der Beteiligung sein können. Hier wurde also bereits versucht, den durch das Vorhandensein lediglich eines einzigen Relationstyps entstehenden Unzulänglichkeiten eingegenzuwirken, indem Kategorien geschaffen wurden, die eine Relation und keinen Begriff bezeichnen. Solche Kategorien können und sollen mit der Existenz eines differenzierten Relationeninventars ja abgeschafft werden.

5.3.2 Intellektuelles Vorgehen

Auf der intellektuellen Ebene bedeutet die Entwicklung eines theaterspezifischen Relationeninventars für das Kategoriensystem nun also zunächst die Betrachtung jeder einzelnen der rund 15.000 vorhandenen Relationen. Die Instanzrelationen und anderen Beziehungen zu Individualbegriffen können dabei rasch behandelt werden.⁶⁵ Dabei werden sie jedoch in der Struktur belassen, um im Anwendungsfall innerhalb der Wikipedia eine Navigation nach Instanzen und Nicht-Instanzen zu ermöglichen und nicht, wie es derzeit der Fall ist, bei einigen Kategorien eine Vermischung von Instanzen und Nicht-Instanzen beizubehalten.⁶⁶

Im Zentrum der Betrachtung stehen naturgemäß die Relationen zwischen Allgemeinbegriffen. Dabei wird versucht, die jeweilige Relation zunächst möglichst genau zu beschreiben und im Verlauf der Arbeit soweit zu abstrahieren, dass wiederkehrende Typen von Relationen identifizierbar sind. Hierbei ist, wie in Abschnitt 4.1 bereits beispielhaft beschrieben, zwischen einer möglichst genauen Beschreibung der einzelnen Relation und einer bestimmten Abstraktion abzuwägen. Wird eine Beziehung so genau beschrieben, dass der so formulierte Relationstyp nur einmal auftritt, so ist er wertlos. Wird wiederum so sehr abstrahiert, dass am Schluss im Extremfall nur die hierarchische und die assoziative Relation übrig bleiben, so ist die Arbeit ebenso wertlos. Es ist also die Mitte zu finden. Dabei sei erneut darauf hingewie-

⁶⁵Es gibt, wie das fertige Inventar zeigt, einerseits Instanzbeziehungen, die hierarchisch sind, und andererseits assoziative Beziehungen, bei denen einer der beiden beteiligten Begriffe einen Individualbegriff bezeichnet.

⁶⁶Beispielsweise enthält die Kategorie „Drama“ 2.360 Instanzbegriffe sowie 29 Begriffe, die keine Instanzen sind.

Theaterbegriff NT Bühnentechnik
 Theaterbegriff NT Musicalbegriff
 Theaterbegriff NT Tanzbegriff
 Theaterbegriff NT Theater_(Architektur)
 Theaterbegriff NT Theaterberuf
 Theaterbegriff NT Theaterfigur
 Theaterbegriff NT Theatergenre
 Theaterbegriff NT Theaterregie
 Theaterbegriff NT a:Ab_durch_die_Mitte
 Theaterbegriff NT a:Abgang_(Theater)
 Theaterbegriff NT a:Akt_(Theater)
 Theaterbegriff NT a:Auftritt_(Theater)
 Theaterbegriff NT a:Aus_der_Rolle_fallen
 Theaterbegriff NT a:Ausstattung_(Theater)
 Theaterbegriff NT a:Beifall
 Theaterbegriff NT a:Beiseitesprechen
 Theaterbegriff NT a:Botenbericht
 Theaterbegriff NT a:Bühnendeutsch
 Theaterbegriff NT a:Bühnenkampf
 Theaterbegriff NT a:Bühnenreifepfung
 Theaterbegriff NT a:Debüt
 Theaterbegriff NT a:Dernière
 Theaterbegriff NT a:Deus_ex_machina
 Theaterbegriff NT a:Einsatz_(Zeitpunkt)
 Theaterbegriff NT a:Einspringer
 Theaterbegriff NT a:En-suite-Spielbetrieb
 Theaterbegriff NT a:Engagement_(Theater)
 Theaterbegriff NT a:Erstaufführung
 Theaterbegriff NT a:Extempore
 Theaterbegriff NT a:Falscher_Abgang
 Theaterbegriff NT a:Fundus_(Sammlung)
 Theaterbegriff NT a:Gammelot
 Theaterbegriff NT a:Hauptrolle
 Theaterbegriff NT a:Hosenrolle
 Theaterbegriff NT a:Kompliment_(Verbeugung)
 Theaterbegriff NT a:Mehrspartentheater
 Theaterbegriff NT a:Nebenrolle
 Theaterbegriff NT a:Premiere
 Theaterbegriff NT a:Repertoiresystem
 Theaterbegriff NT a:Schmierentheater
 Theaterbegriff NT a:Semi-Stagione
 Theaterbegriff NT a:Spielzeit_(Theater)
 Theaterbegriff NT a:Stagionesystem
 Theaterbegriff NT a:Stückvertrag
 Theaterbegriff NT a:Szene_(Theater)
 Theaterbegriff NT a:Travestie
 Theaterbegriff NT a:Umbesetzung
 Theaterbegriff NT a:Uraufführung
 Theaterbegriff NT a:Verfremdungseffekt
 Theaterbegriff NT a:Vierspartenhaus
 Theaterbegriff NT a:Vierte_Wand
 Theaterbegriff NT a:Vorpremiere
 Theaterbegriff NT a:Wiederaufnahme_(Theater)

Abbildung 5: Direkt untergeordnete Kategorien und Artikel der Kategorie „Theaterbegriff“ im bisherigen Stand

sen, dass das zu entwickelnde Inventar zunächst einmal theaterspezifisch sein soll und also nicht an jeder Stelle soweit abstrahiert werden muss, dass ein Relationstyp vollkommen themenübergreifend wird. Dies soll ja erst im zweiten Arbeitsfeld der Bearbeitung des Bereichs Jagd erfolgen.

Zur Definition der Begriffe, die ja für die Formulierung der Relationstypen unumgänglich ist, wurden lediglich die jeweiligen Artikel der Wikipedia herangezogen. Dies ist zu begründen, indem davon ausgegangen wird, dass ein Artikel nur gemäß seines Textes und also der im Text formulierten Bedeutung seines Namens ins Kategoriensystem eingefügt wurde. Ein Relationeninventar für die Wikipedia muss ja naturgemäß von den durch die Artikelaufsteller und in der Kategorisierung Tätigen intendierten Begriffsbedeutungen geleitet sein. Andere Nachschlagewerke aus dem Theaterbereich wurden daher nicht herangezogen.

5.3.3 Technisches Vorgehen

Die bereits genannten Dumps der Wikipedia werden jeweils im Abstand von einigen Wochen erzeugt, und diejenigen der deutschsprachigen Wikipedia in der Verzeichnisstruktur beginnend bei der URL <http://dumps.wikimedia.org/dewiki/> veröffentlicht.⁶⁷ Der dieser Arbeit zugrunde liegende Dump vom 3. Juni 2010 ist unterhalb des Verzeichnisses <http://dumps.wikimedia.org/dewiki/20100603/> abgelegt. Dabei gibt es keine einzelne große Datei, die sämtliche Daten der jeweiligen Sprachversion enthält, sondern verschiedene Dateien getrennt nach den verschiedenen technischen Bereichen der Wikipedia. Für diese Arbeit wurden die folgenden Dateien benötigt:

`page.sql.gz` enthält formale Angaben zu allen „pages“ der Wikipedia: Identifikationsnummer (ID), Typ der Seite („Namespace“), Name, Größe in Bytes und so weiter. Dies beinhaltet sowohl die Artikel, die zum Namespace „0“ gehören, sowie die Kategorien, die zum Namespace „14“ gehören. Nicht enthalten sind die Inhalte der „pages“, also die Artikeltexte beziehungsweise Angaben über die einer Kategorie untergeordneten „pages“.

`categorylinks.sql.gz` gibt anhand der Identifikationsnummern die einer Kategorie untergeordneten Artikel und Kategorien an.

Da die der Wikipedia zugrunde liegende Software MediaWiki anhand einer relationalen Datenbank in der Datenbanksprache SQL arbeitet, liegen die genannten Dumps als SQL-Dateien vor, dabei sind sie zudem mit dem Programm „gzip“ komprimiert. Die SQL-Dateien enthalten Befehle zur Definition einer sogenannten Tabelle sowie zum Befüllen dieser Tabelle mit Inhalten. Die Tabelle `page`, die in der Datei `page.sql.gz` definiert wird, enthält beispielsweise die Spalten `page_id` (Spalte 1,

⁶⁷Vgl. [Wikimedia 2010].

für die Identifikationsnummer), `page_namespace` (Spalte 2, für die Angabe des Namespaces), `page_title` (Spalte 3, für die Angabe des Namens) und so weiter. Nach der Definition der Spalten folgen Befehle, die die Tabelle mit Inhalten füllen, also namentlich die konkreten Informationen. Der Befehl

```
INSERT INTO 'page' VALUES (1,0,'Alan_Smithee','',156,0,0,
0.0864337124735431,'20100309161334',71650845,4341,0)
```

gibt etwa an, dass die Seite mit `page_id` 1 zu `page_namespace` 0 gehört und den Namen `Alan_Smithee` hat. Hier wird also, erkennbar am Namespace 0, der Artikel „Alan Smithee“ formal in die Datenbank aufgenommen, mithin also angegeben, dass er in der deutschsprachigen Wikipedia existiert. Die weiteren Angaben hinter dem Namen werden für diese Arbeit nicht benötigt. Sie werden zusammen mit allen anderen Tabellen (`categorylinks` und so weiter) in Anleitung zu MediaWiki unter [MediaWiki 2010] beschrieben.

Da für diese Arbeit jedoch keine relationale Datenbank benötigt wurde, wurden die SQL-Befehle jedoch nicht ausgeführt, sondern die Dumps nur zur Extraktion der benötigten Informationen beziehungsweise Daten verwendet. Das Herausziehen der Daten sowie deren Darstellung in geeigneten Formaten erfolgte mittels mehrerer Skripte im Format der Unix-Shell „Bash“ („Bourne-again shell“), die vom Autoren dieser Arbeit geschrieben wurden. Sie sind im Anhang dieser Arbeit abgedruckt sowie auf der beigelegten CD-ROM im Verzeichnis `skripte/` enthalten.⁶⁸

Im Folgenden werden die Skripte genannt sowie ihre Aufgaben beschrieben:⁶⁹

umarbeitung-dumps Extraktion ausschließlich der benötigten Daten aus den Dumps.

Dabei wird eine Datei `page.gz` erzeugt, die pro Zeile die ID, den Namespace und den Namen einer „page“ enthält sowie eine Angabe darüber, ob es sich um eine Weiterleitung handelt, außerdem eine Datei `categorylinks.gz`, die pro Zeile die Angabe der ID einer „page“ enthält sowie den Namen einer Kategorie, der sie zugeordnet ist.

browse-down Sucht, ausgehend von dem Namen einer Kategorie, die untergeordneten Kategorien und Artikel. Dabei werden zunächst anhand der Datei `categorylinks.gz` alle der Kategorie untergeordneten `page_ids` bestimmt, daraufhin werden anhand der Datei `page.gz` die Namen zu den IDs sowie ihr Namespace ermittelt. Lautet der Namespace „14“ und handelt es sich bei der

⁶⁸Da der Autor dieser Arbeit kein Programmierer, sondern Bibliothekar ist, ist es wahrscheinlich, dass die Skripte nicht effizient sind und dass es algorithmisch bessere Lösungen zur Erfüllung der Aufgaben gibt. Solche Einwände sind für diese Arbeit jedoch nicht von Belang. Es war beim Verfassen der Skripte lediglich wichtig, *dass* sie ihre Aufgaben erfüllen.

⁶⁹Eine weitergehende Beschreibung der Skripte sowie der einzelnen Befehle erfolgt an dieser Stelle nicht. Bei Interesse sei zum besseren Verständnis auf Literatur zu „bash“ verwiesen, beispielsweise [Newham 2005].

untergeordneten „page“ also ihrerseits um eine Kategorie, so ruft sich das Skript mit der ermittelten Kategorienbezeichnung selbst auf. Das Skript gibt die ermittelten Angaben zeilenweise aus, wobei jede Zeile dem Format

Kategorie NT untergeordnete_„page“

folgt. Ist das Untergeordnete eine Kategorie, so wird lediglich ihr Name angegeben, ist es ein Artikel, so wird der Name mit dem Präfix „a:“ angegeben.⁷⁰ Die Reihenfolge der ausgegebenen Zeilen folgt den aufsteigenden IDs und keiner alphabetischen oder gar inhaltlichen Ordnung. Dabei folgt, wenn das Untergeordnete eine Kategorie ist, durch den Selbstaufwurf des Skriptes zunächst das Untergeordnete dieser Kategorie, erst danach wird zur übergeordneten Kategorie zurückgekehrt. So wird bereits die im Kategoriensystem konzeptionell suggerierte Baumstruktur erkennbar.

browse-up Geht entgegengesetzt zu **browse-down** von einem Artikelnamen aus und sucht seine übergeordneten Kategorien. Dieses Skript wurde in einem frühen Stadium dieser Arbeit verfasst und fand später keine Verwendung.

tree-top-down-plain Erzeugt anhand der Ausgabe von **browse-down** eine Textdatei mit Zeilen im Format

Kategorie UB untergeordnete_„page“

in alphabetischer Reihenfolge. Die von diesem Skript erzeugte Datei folgt ebenfalls nach wie vor der im Kategoriensystem suggerierten Baumstruktur, wobei alles Untergeordnete im Gegensatz zur Ausgabe von **browse-down** jedoch alphabetisch sortiert ist. Die deutsche Bezeichnung „UB“ erzeugt naturgemäß ebenfalls die in Fußnote 70 dargestellte Problematik, zeigt jedoch an, dass an dieser Stelle eine Hinwendung zur deutschsprachigen Wikipedia und zu einem deutschsprachigen Relationeninventar erfolgte. Die Ausgabe dieses Skriptes wurde zur Entwicklung des Relationeninventars genutzt, indem in jeder Zeile die Bezeichnung „UB“ durch eine Bezeichnung der wirklichen Beziehung zwischen dem links- und dem rechtsstehenden Begriff ersetzt wurde.

⁷⁰ Die Bezeichnung „NT“ („narrow term“) suggeriert, dass es sich bei jeder solchen Zeile um die Angabe der Beziehung zwischen dem links davon und dem rechts davon stehenden Begriff handelt, wobei der rechtsstehende dem linksstehenden untergeordnet sei. Die Bezeichnung „NT“ erfolgte auch an dieser Stelle jedoch lediglich gemäß dem im Kategoriensystem verfolgten Konzept der Über- und Unterordnung von Kategorien. Dass eine solche Über- und Unterordnung der durch die Kategorien- beziehungsweise Artikelnamen bezeichneten Begriffe nicht immer gegeben ist, wurde in dieser Arbeit bereits erläutert. Die englische Bezeichnung „NT“ wurde an dieser Stelle der deutschen Übersetzung „UB“ (Unterbegriff) vorgezogen, da das Skript so auch in einem Dump der englischsprachigen Wikipedia Anwendung finden kann, wenn auch eine solche Benutzung nicht mehr Teil dieser Arbeit ist.

tree-top-down-html Erzeugt anhand der Ausgabe von **browse-down** eine HTML-Datei, die die Strukturen unterhalb der Ausgangskategorie anhand verschachtelter Listen darstellt. Dabei werden auch Hyperlinks zu den jeweiligen „pages“ der deutschsprachigen Wikipedia eingesetzt. Diese Ausgabe diente dem Autoren dieser Arbeit zur Gewinnung einer Übersicht über den jeweiligen Teil des Kategoriensystems.

Für die Arbeit im Gegenstandsbereich Theater wurden also in der Shell die folgenden Befehle ausgeführt:

```
$ umarbeitung-dumps page
$ umarbeitung-dumps categorylinks
$ browse-down Theater > Theater-down
$ tree-top-down-html > Theater-down.html
$ tree-top-down-plain > Theater-down-sortiert.txt
```

Die erzeugte Datei **Theater-down-sortiert.txt** wurde daraufhin in einem Texteditor aufgerufen, und es wurde in sämtlichen Zeilen der provisorische Beziehungstyp „UB“ durch den jeweiligen wirklichen Beziehungstypen ersetzt. So erfolgte die Entwicklung des Relationeninventars. Die Datei mit den wirklichen Beziehungstypen wurde **Theater-bearbeitet.txt** genannt.

Die von den Skripten erzeugten Dateien **Theater-down**, **Theater-down.html** und **Theater-down-sortiert.txt** sowie die manuell bearbeitete Datei **Theater-bearbeitet.txt** sind sämtlich auf der dieser Arbeit beigelegten CD-ROM im Verzeichnis **theater/** enthalten.

Analog wurde im Gegenstandsbereich Jagd verfahren, vgl. hierzu Abschnitt 5.4 dieser Arbeit.

Eine über die zeilenweise Darstellung in **Theater-bearbeitet.txt** hinausgehende Visualisierung der neuen Strukturen erfolgte nur exemplarisch. Das technische Vorgehen zur Visualisierung wird in Abschnitt 5.3.5 beschrieben.

5.3.4 Das neue Relationeninventar

Wie im vorigen Abschnitt beschrieben, erfolgte die Entwicklung des Relationeninventars, indem die Datei **Theater-down-sortiert.txt** zeilenweise bearbeitet wurde. Dabei wurden die am linken und rechten Ende jeder Zeile stehenden Kategorien- beziehungsweise Artikelnamen, wenn nötig, in Deskriptoren überführt, indem sie über die gesamte Datei vereinheitlicht wurden – dies betraf in der Regel das Zurückführen von Pluralformen auf ihre Singularform –, und die in der Mitte stehende provisorische Beziehungsbezeichnung „UB“ ersetzt durch eine Beziehungsbezeichnung des neuen,

hierarchisch gegliederten Inventars, das ja eben durch diese Ersetzungsarbeit entstand. Die Zeilen der sich so entwickelnden Datei `Theater-bearbeitet.txt` haben also die folgende Form:

Begriff_1 Relationstyp Begriff_2

Zur besseren Handhabbarkeit durch Skripte und andere Software dienen Leerzeichen als Trennung zwischen den drei Teilen eines solchen Tripels. Erscheinen in den beiden Begriffen Leerzeichen, so sind sie, wie bereits in den in SQL formulierten Ausgangsdateien, durch Unterstriche („_“) ersetzt. In den Bezeichnungen der Relationstypen erscheinen ebenso Unterstriche, hier dienen diese jedoch zur Trennung der Hierarchiestufen im Relationeninventar: Jeder Unterstrich bedeutet eine Hierarchiestufe.

Zwei beispielhafte Zeilen lauten so:

```
Theater UB_generisch_Form_Sparte Drama
Drama UB_generisch_Form_Stoff Komödie
```

Sie bedeuten, dass das Drama eine Form des Theaters nach dem Aspekt der Theatersparte ist, und dass die Komödie eine Form des Dramas nach dem Aspekt des behandelten Stoffes ist.

Sofern eine Beziehung auf mehreren Aspekten gründet, können zwischen zwei Begriffen mehrere Relationstypen angegeben werden. Dies erfolgt, indem die zwei Begriffe auf einer zweiten oder mehreren Zeilen wiederholt werden:

```
Komödie UB_generisch_Beteiligte_Publikum Schwank
Komödie UB_generisch_Form_Stoff Schwank
```

So wird ausgedrückt, dass ein Schwank eine Komödie ist, die sich an ein bestimmtes Publikum richtet und außerdem bestimmten Stoff behandelt.

Die Option der Angabe mehrerer Beziehungstypen zwischen zwei Begriffen ermöglicht eine differenziertere Suche: Ein Nutzer des Vokabulars kann etwa nach allen Unterbegriffen des Begriffs „Theater“ suchen, die auf der Form nach dem Aspekt des behandelten Stoffes gründen und außerdem einen lokalen Bezug zu einem geographischen Ort haben, also beispielsweise typisch für ein bestimmtes Land sind. Um dieses zu ermöglichen, sollen in einer Anwendung bei Navigation und Suche Relationstypen also mit Boole’schen Operatoren verknüpft werden können.

Von der Neurelationierung ausgenommen wurden diejenigen Zeilen, deren linke beziehungsweise rechte Seite keinen Begriff im Sinne von [Wersig 1978] bezeichnet, also etwa:

```
Theater -- Theatergenre
```

Um Rückschlüsse auf die ursprüngliche Struktur des Vokabulars zu ermöglichen, wurden solche Zeilen jedoch nicht gelöscht, sondern, wie zu sehen ist, anstatt eines Relationstypen durch „--“ markiert. Im dargestellten Fall wurden in den darauffolgenden Zeilen die linken Seiten, die ja „Theatergenre“ lauteten, durch „Theater“ ersetzt. Dies bedeutet, dass die zwischengeordnete Kategorie „Theatergenre“, die ja nur in Ermangelung differenzierter Relationstypen eingeführt wurde, durch eine direkte Relationierung des Begriffs „Theater“ und der Genrebegriffe ersetzt wurde.

Die vorige Struktur lautete also beispielsweise:⁷¹

```
Theater UB Theatergenre
Theatergenre UB a:Absurdes_Theater
Theatergenre UB a:Ausstattungsstück
Theatergenre UB a:Bauerntheater
```

Die neue Struktur lautet hingegen:

```
Theaterbegriff -- Theatergenre
Theater UB_generisch_Form_Aufbau a:Absurdes_Theater
Theater UB_generisch_Form_Stoff a:Absurdes_Theater
Theater UB_generisch_Zeit_historisch a:Absurdes_Theater
Theater UB_generisch_Form_Darstellung a:Ausstattungsstück
Theater UB_generisch_Form_Stoff a:Bauerntheater
Theater UB_generisch_Beteiligte_Darstellende a:Bauerntheater
Theater UB_generisch_Beteiligte_Publikum a:Bauerntheater
```

Wie zu sehen ist, wurde jedoch die Trennung von Kategorien und Artikeln beibehalten, indem diejenigen Deskriptoren, die aus Artikelnamen entstanden, weiterhin mit dem Präfix „a:“ markiert werden. So ist bei einer Verwendung des neuen Inventars innerhalb der Wikipedia weiterhin eine nach Kategorien und Artikeln getrennte Navigation möglich.

Am Beginn der Relationierungsarbeit begann die Formulierung der neuen Relationstypen naturgemäß, umgangssprachlich ausgedrückt, ins Blaue hinein. Erst im Verlauf der Arbeit gelang es, einen vollständigen Überblick über die Struktur des Vokabulars zu erhalten und wiederkehrende Muster zu erkennen, die in wiederkehrenden Relationstypen mündeten. So musste die zeilenweise Arbeit mehrfach unterbrochen und zu bereits bearbeiteten Zeilen zurückgekehrt werden, um vorher getätigte Formulierungen der Relationstypen zu verändern.

Das fertige Relationeninventar wird mit kurzen Definitionen der einzelnen Beziehungstypen sowie einer Tabelle zur Häufigkeitsverteilung der Relationstypen in

⁷¹Gezeigt wird an dieser Stelle nur ein kleiner Teil der der Kategorie „Theatergenre“ untergeordneten Kategorien und Artikel. Im Vokabular sind es weit mehr.

Anhang A dargestellt. Wie bereits beschrieben, ist es auf den Gegenstandsbereich des Theaters zugeschnitten.

In der generischen hierarchischen Relation zeigen sich Verfeinerungen nach der Form, nach den Beteiligten sowie nach örtlichen und zeitlichen Bezügen. Die formalen Aspekte sollen an dieser Stelle näher erläutert werden: Der Aspekt der Sparte bezieht sich lediglich auf die vier klassischen Sparten des Theaters (Sprechtheater, Musiktheater, Tanztheater sowie Figurentheater). Der Aspekt der Methode bezieht sich auf die grundlegend verwendeten Methoden, sofern sie nicht die vier klassischen Sparten betreffen, beziehungsweise wenn Elemente einer Sparte methodisch in eine andere übernommen werden. Mit dem Aspekt der Darstellung ist eine Darstellungsform innerhalb einer Sparte gemeint, mit dem Aspekt des Ortes ein Darstellungsort. Der Begriff Aufbau bezeichnet in diesem Zusammenhang den formalen Aufbau, der Begriff Stoff den behandelten Stoff (das Inhaltliche).

Die assoziative Relation wurde aus Gründen der sprachlichen Einfachheit und Allgemeinverständlichkeit mit der aus dem Umfeld von Thesauri stammenden Bezeichnung des „verwandten Begriffs“ versehen. Analog zum Verfahren bei assoziativen Relationen in Thesauri, wurde auch hier keine Richtung dieser Beziehungstypen formal festgehalten, obgleich die meisten dieser Typen naturgemäß gerichtet sind.⁷² Hierauf wurde verzichtet, da das Relationeninventar und damit relationierte Vokabulare, wie bereits beschrieben, durch Menschen verwendet werden sollen, die die Richtung einer Beziehung selbst erkennen können, wie es ja auch bei den verwandten Begriffen in Thesauri geschieht. Die Einführung jeweils eines Beziehungspaares, das zwei entgegengesetzte Beziehungen ausdrückte, würde das Inventar unnötig komplexer machen. Soll das Inventar jedoch in Zukunft einmal zur Inferenz durch Maschinen eingesetzt werden, so muss naturgemäß festgehalten werden, ob eine Beziehung aus den assoziativen Typen symmetrisch ist, und, falls nicht, die Richtung der Beziehung.

Innerhalb der formulierten assoziativen Relationen fällt zunächst die bereits in Abschnitt 4.1 vorgesehene theaterspezifische Dreiteilung von Erschaffung im kreativen Prozess, Gestaltung und Darstellung auf. Darüber hinaus wird eine Beziehung nach dem Aspekt der Betrachtung außerhalb dieser Dreiteilung der aktiven Beteiligung eingeführt, mithin also eine Beziehung, die eine passive Beteiligung am Theaterschaffungsprozess anzeigt. Daneben gibt es eine Beziehung, die eine sonstige Beteiligung anzeigt, die nicht durch die jetzt genannten vier Typen abgedeckt wird. Innerhalb dieser Relationstypen kann eine hierarchische Ordnung erkannt werden, die im entwickelten Relationeninventar jedoch, im Gegensatz zu den ersten Überlegungen in Abschnitt 4.1, noch nicht abgebildet wird. Der Grund hierfür liegt in der bereits

⁷²Wenn **Begriff_2** beispielweise ein Mittel beschreibt, das für die mit **Begriff_1** beschriebene Tätigkeit verwendet wird, dann kann diese Beziehung nicht außerdem umgekehrt bestehen. Sie ist also nicht symmetrisch.

angedeuteten Schwammigkeit des Begriffs der Beteiligung. Fasst man ihn weit, so bedeutet jegliche assoziative Beziehung eine wie auch immer geartete Beteiligung. Fasst man ihn eng und bezieht ihn auf aktives Tun, so würde er nur Personen und Personengruppen umfassen. Im Sinne der Allgemeinverständlichkeit wurde im theaterspezifischen Inventar also auf eine Einführung einer Hierarchiestufe „Beteiligung“ verzichtet, wenngleich nun der Beziehungstyp der „sonstigen Beteiligung“ einen unglücklichen Eindruck macht.

Die Beziehung der „Hervorbringung“ ist leider unscharf, schwammig und damit unglücklich. Sie umfasst jegliche Beziehung zwischen etwas, das entsteht, und dem Hervorbringenden, die nicht direkt innerhalb des Theaterschaffungsprozesses von Erschaffung, Gestaltung und Darstellung (in der Regel also Aufführung) steht. Dies schließt Beziehungen zwischen Produzent und Produkt ein, geht jedoch über sächliche Produkte hinaus. Hier ergibt sich eine Unschärfe betreffend die vorher beschriebenen Beziehungen in der „Beteiligung“.

Auf die weiteren Relationstypen soll an dieser Stelle nicht näher eingegangen und stattdessen auf den Anhang A verwiesen werden. In der auf der beigelegten CD-ROM enthaltenen Datei `Theater-bearbeitet.txt` finden sich naturgemäß sämtliche relationierten Begriffe und damit Beispiele für alle Relationstypen.

5.3.5 Exemplarische Darstellung der neuen Strukturen mit Graphviz

Damit das entwickelte Relationeninventar und die neuen Strukturen des Theatervokabulars angewendet werden können, müssen hierfür die technischen Voraussetzungen in der eingesetzten Software gegeben sein. Soll das Inventar, wie vorgeschlagen, direkt in der Wikipedia eingesetzt werden, so bietet es sich an, die der Wikipedia zugrunde liegende Software MediaWiki sowie das Tool Catgraph so zu erweitern, dass sie eine beliebige Zahl an Relationstypen verarbeiten können, besser noch, dass sie darüber hinaus mit hierarchisch geordneten Relationstypen umgehen können.

Solche Erweiterungen würden jedoch den Rahmen dieser Arbeit sprengen. Dennoch erscheint es an dieser Stelle wünschenswert, die neuen Vokabularstrukturen bereits jetzt nicht nur innerhalb der Datei `Theater-bearbeitet.txt` zur Verfügung zu stellen, sondern zumindest auch eine grundlegende und simple graphische Visualisierung anzubieten, um sie dem Leser dieser Arbeit anschaulicher zu machen.

Dies kann mit dem Programmpaket Graphviz geleistet werden, auf dem auch Catgraph basiert.⁷³ Graphviz beinhaltet mehrere Programme, die das Darstellen von Graphen ermöglichen, deren Knoten und Kanten mit einer spezifischen Beschreibungssprache formuliert werden. Dabei arbeitet jedes Programm mit einem anderen

⁷³Dokumentation zu Graphviz sowie die Möglichkeit des Herunterladens des Programmpakets bietet die entsprechende Website [Graphviz 2010].

Darstellungsalgorithmus. Für die vorliegenden Vokabularstrukturen bietet sich nicht etwa eine Baumstruktur, sondern ein netz- beziehungsweise wolkenartiges Layout an. Dies leistet das zu Graphviz gehörende Programm `fdp`.

`fdp` benötigt eine Eingabedatei mit der Graphenbeschreibung im Graphviz-Format. Eine solche Datei kann aus dem Datenformat von `Theater-bearbeitet.txt` erzeugt werden. Dabei ist es zweckmäßig, nicht sämtliche im Vokabular vorkommenden Begriffe und Relationen in einem Graphen darzustellen, sondern bestimmte Begriffe und Relationen auszuwählen.

Zur Umwandlung der ausgewählten Ausschnitte aus `Theater-bearbeitet.txt` ins Graphviz-Format hat der Autor dieser Arbeit zwei Skripte für die Unix-Shell „bash“ geschrieben, die ebenfalls im Anhang D dieser Arbeit abgedruckt und auf der beigelegten CD-ROM im Verzeichnis `skripte/` enthalten sind.

Das Skript `visualisierung-or` erlaubt die Angabe eines Ausgangsbegriffs und einem oder mehrerer Relationstypen und zeigt diejenigen der vom Ausgangsbegriff ausgehenden Relationen an, die mindestens einem der angegebenen Typen entsprechen.⁷⁴ Es ist die Angabe von bis zu neun Relationstypen möglich. Wird nur ein Ausgangsbegriff, jedoch kein Relationstyp angegeben, so werden maximal neun aller Zusammenhang mit dem Ausgangsbegriff auftretenden Relationstypen dargestellt. Beziehungen zu Individualbegriffen werden aufgrund ihrer in der Regel hohen Zahl, die den Graphen unübersichtlich machte, dabei jedoch außen vor gelassen.⁷⁵

`visualisierung-or` gibt Quelltext im Graphviz-Format aus, der daraufhin mit `fdp` verarbeitet werden kann. Hierarchische Beziehungen werden dabei anhand von Pfeilen symbolisiert, bei den assoziativen Beziehungen, bei denen, wie erläutert, keine Richtung formal festgehalten wird, erscheint naturgemäß kein Pfeil. Die Kanten erscheinen entsprechend den bis zu neun Relationstypen in bis zu neun verschiedenen Farben, die in einer Legende erläutert werden.⁷⁶ Wird eine Relation durch zwei oder mehr Relationstypen beschrieben, so erscheinen auch mehrere Kanten zwischen den Begriffen. Ein Begriff, der aus einem Kategoriennamen gebildet wurde, erscheint umgeben von einer Ellipse, ein solcher, der aus einem Artikelnamen gebildet wurde, erscheint umgeben von einem Rechteck. Der Begriff, von dem ausgegangen wurde, ist zur raschen Identifikation von einem Oktagon mit doppelten Linien umgeben.

Das Skript `visualisierung-and` erlaubt die Angabe eines Ausgangsbegriffs und einem oder mehrerer Relationstypen, wobei nur diejenigen der Zielbegriffe angezeigt werden, die durch *sämtliche* der angegebenen Relationstypen miteinander verbunden

⁷⁴Das „or“ im Skriptnamen bedeutet hier den Boole'schen Operator „OR“ und heißt eben, dass alle Relationen angezeigt werden, die *mindestens einem* der angegebenen Relationstypen entsprechen.

⁷⁵Dies schließt sowohl die hierarchischen Instanzrelationen, als auch die assoziativen Beziehungen zu Instanzen ein.

⁷⁶Dass nur maximal neun Relationstypen verarbeitet werden können, rührt daher, dass die verwendete Farbpalette nur neun Farben enthält. Technisch ist bei der Verwendung einer anderen Farbpalette jedoch auch das Darstellen einer höheren Zahl von Relationstypen möglich.

sind.⁷⁷ Darüber hinaus arbeitet dieses Skript wie `visualisierung-or`.

Es folgen einige beispielhafte Anwendungen der Skripte in einer Unix-Shell:

1. `$ visualisierung-or Theater-bearbeitet.txt Theaterregie | fdp -Tpdf`
Erzeugt eine PDF-Datei mit allen Begriffen, zu denen von Begriff Theaterregie eine Beziehung besteht.
2. `$ visualisierung-or Theater-bearbeitet.txt Theater \`
`UB_generisch_Form_Sparte | fdp -Tpdf`
Erzeugt eine PDF-Datei mit denjenigen Unterbegriffen des Begriffs Theater, die eine Theatersparte beschreiben.
3. `$ visualisierung-or Theater-bearbeitet.txt Theater \`
`UB_generisch_Form_Stoff | fdp -Tpdf`
Erzeugt eine PDF-Datei mit denjenigen Unterbegriffen des Begriffs Theater, die Genres nach dem behandelten Stoff beschreiben.
4. `$ visualisierung-or Theater-bearbeitet.txt Theater \`
`UB_generisch_Form_Aufbau UB_generisch_Form_Stoff | fdp -Tpdf`
Erzeugt eine PDF-Datei mit denjenigen Unterbegriffen des Begriffs Theater, die Genres nach dem formalen Aufbau und/oder dem behandeltem Stoff beschreiben.
5. `$ visualisierung-and Theater-bearbeitet.txt Theater \`
`UB_generisch_Form_Aufbau UB_generisch_Form_Stoff | fdp -Tpdf`
Erzeugt eine PDF-Datei mit denjenigen Unterbegriffen des Begriffs Theater, die Genres nach dem formalen Aufbau *und* dem behandeltem Stoff beschreiben.
6. `$ visualisierung-and Theater-bearbeitet.txt Theater \`
`UB_generisch_Form_Stoff UB_generisch_Zeit_historisch | fdp -Tpdf`
Erzeugt eine PDF-Datei mit denjenigen Unterbegriffen des Begriffs Theater, deren Unterordnung auf dem Aspekt des behandelten Stoff *und* einem historischen Bezug gründet.

Die Ausgabe von Befehl 1 ist in Abbildung 6 dargestellt, die Ausgabe von Befehl 6 in Abbildung 7. Die von den Skripten erzeugten Quelltexte sind in Anhang C.1 abgedruckt. Weitere Abbildungen ohne die dazugehörigen Quelltexte enthält Anhang C.2 dieser Arbeit.

⁷⁷Das „and“ im Skriptnamen bedeutet hier den Boole'schen Operator „AND“ und heißt eben, dass in diesem Fall nur die Begriffe angezeigt werden, die durch *sämtliche* der angegebenen Relationstypen miteinander verbunden sind.

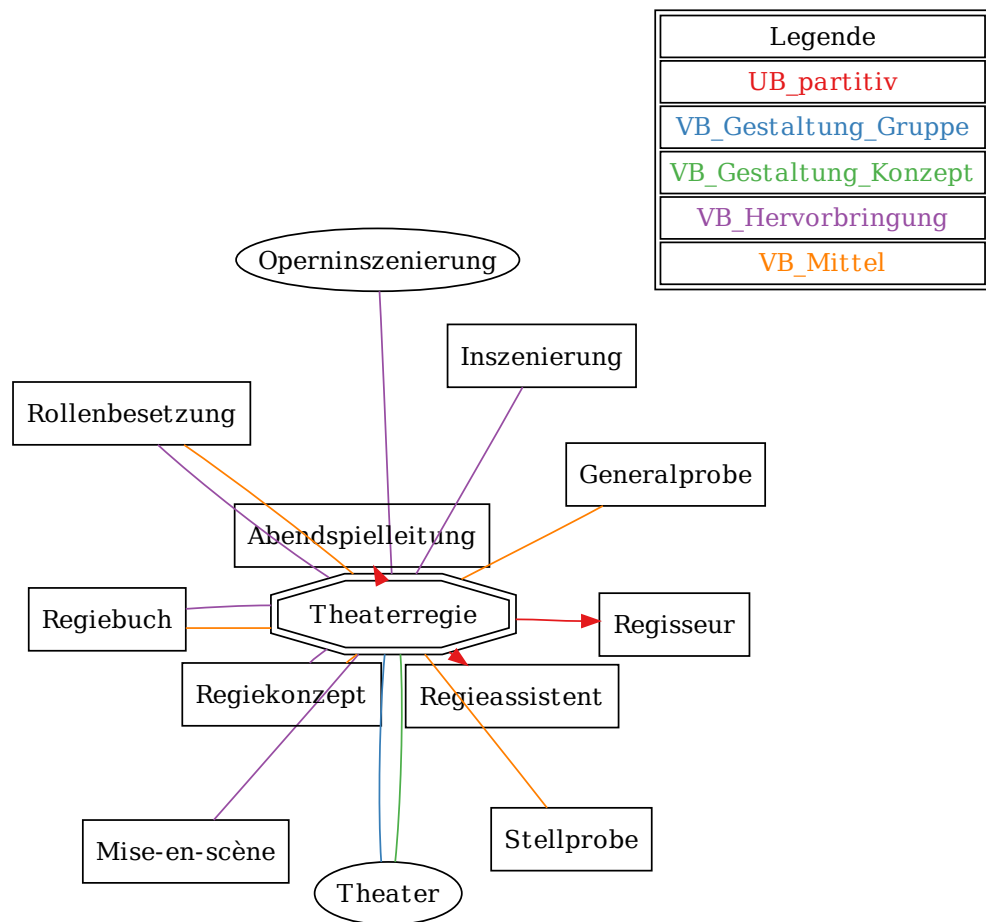


Abbildung 6: Das semantische Umfeld des Begriffs Theaterregie in den neuen Strukturen

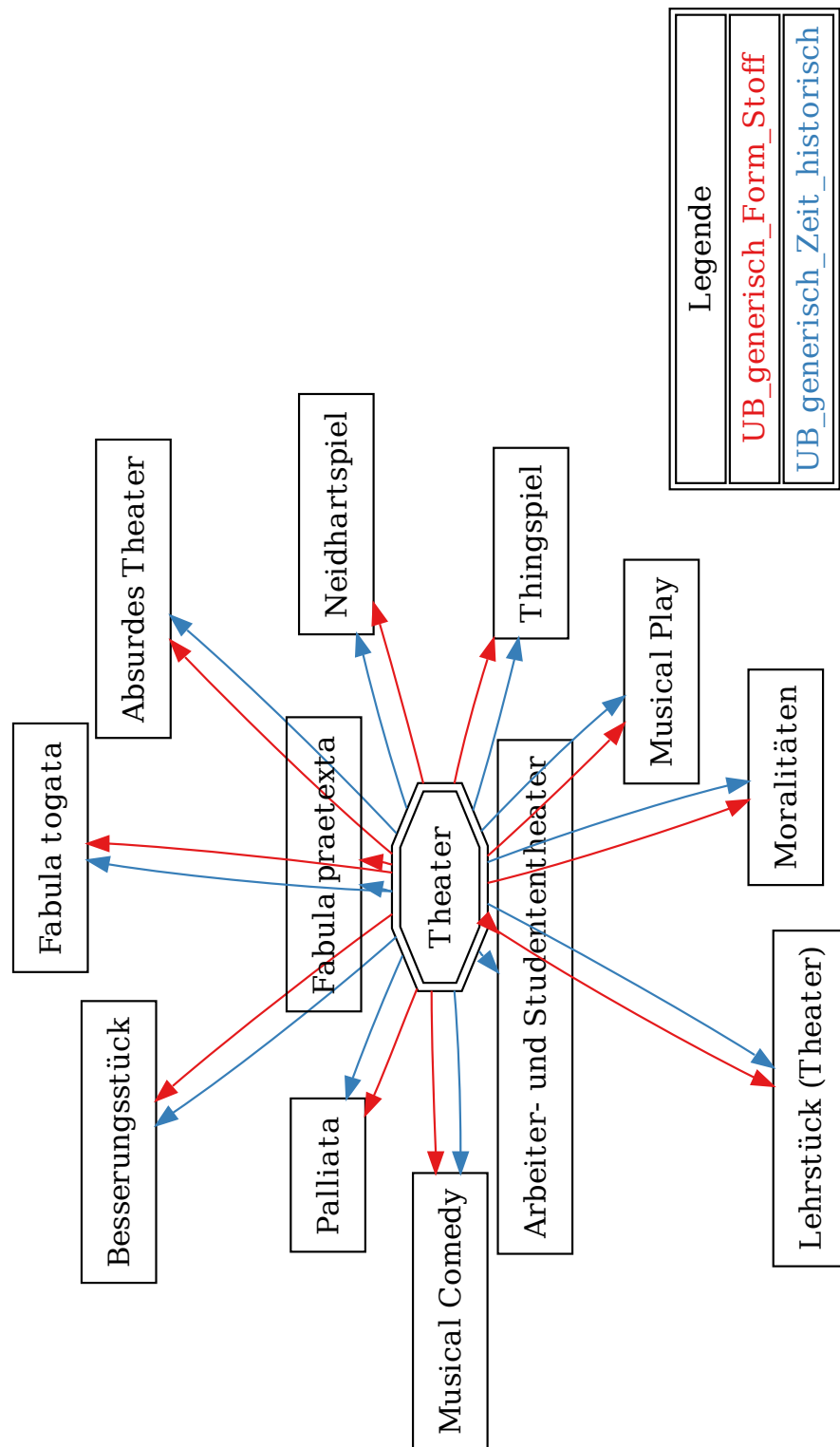


Abbildung 7: Die Unterbegriffe des Begriffs Theater, deren Unterordnung auf dem Aspekt des behandelten Stoffs *und* einem historischen Bezug gründet, in den neuen Strukturen

Wie bereits angedeutet, stellen die beiden Skripte keine Schnittstelle für Endnutzer dar, sondern dienen lediglich der exemplarischen Darstellung der neuen Strukturen zum jetzigen Zeitpunkt. So ist es beispielsweise mit den Skripten allein nicht möglich, Relationen über mehr als zwei Hierarchiestufen darzustellen. Dies kann nur geschehen, in dem von den Skripten erzeugte Graphviz-Quelltexte manuell zusammengefügt werden. Abbildung 8 zeigt hierfür ein Beispiel. Ebenso ist keine automatische Expansion von Relationstypen möglich: Die Angabe von `UB_generisch_Form` führt etwa nicht zu, dass alle Relationstypen auf den darunterliegenden Ebenen (nach den Aspekten Aufbau, Stoff und so weiter) einbezogen werden. Dies muss manuell geschehen.

5.4 Anwendung des Inventars auf den Gegenstandsbereich Jagd

Da das für den Gegenstandsbereich Theater entwickelte Inventar stark auf diesen Bereich zugeschnitten ist, wurde nach dessen Entwicklung versucht, es auf einen anderen Gegenstandsbereich anzuwenden, mit dem vom Theater keinerlei Überschneidungen bestehen. So können Rückschlüsse darauf gezogen werden, ob die Entwicklung eines themenübergreifenden Relationeninventars überhaupt sinnvoll ist, beziehungsweise welche der entwickelten Beziehungstypen es sind, die offensichtlich gegenstandsübergreifend sind und demzufolge möglicherweise für ein themenübergreifendes Inventar eingesetzt werden können.

Das technische Vorgehen war analog zu dem im Theaterbereich. Ausgegangen wurde hier von der Kategorie „Jagd“ der deutschsprachigen Wikipedia. Dieser Bereich des Kategoriensystems ist weit kleiner als der theaterspezifische, er umfasst 846 Kategorien und Artikel, davon 470 Instanzen. Die entsprechenden Dateien `Jagd-down`, `Jagd-down.html`, `Jagd-down-sortiert.txt` und `Jagd-bearbeitet.txt` sind ebenfalls auf der dieser Arbeit beigelegten CD-ROM im Verzeichnis `jagd/` enthalten.

Auch bei diesem Teil der Relationierungsarbeit wurde die Datei `Jagd-down-sortiert.txt` zeilenweise bearbeitet, es wurde jedoch versucht, jede Beziehung direkt durch einen Typen aus dem theaterspezifischen Inventar zu beschreiben, wobei aber, wenn nötig, auch Spezifika der Jagd eingearbeitet wurden. So entstand ein jagdspezifisches Inventar, das in seiner Struktur jedoch dem vorher entwickelten Theaterinventar folgt. Es ist zusammen mit einer Häufigkeitsverteilung im Anhang B dargestellt. Zwei beispielhafte Visualisierungen, die mit den im vorigen Abschnitt beschriebenen Skripten erzeugt wurden, sind im Anhang C.2 abgedruckt. Durch Zusammenführen beider Inventare ergibt sich nun die Möglichkeit der Entwicklung eines themenübergreifenden Relationeninventars.

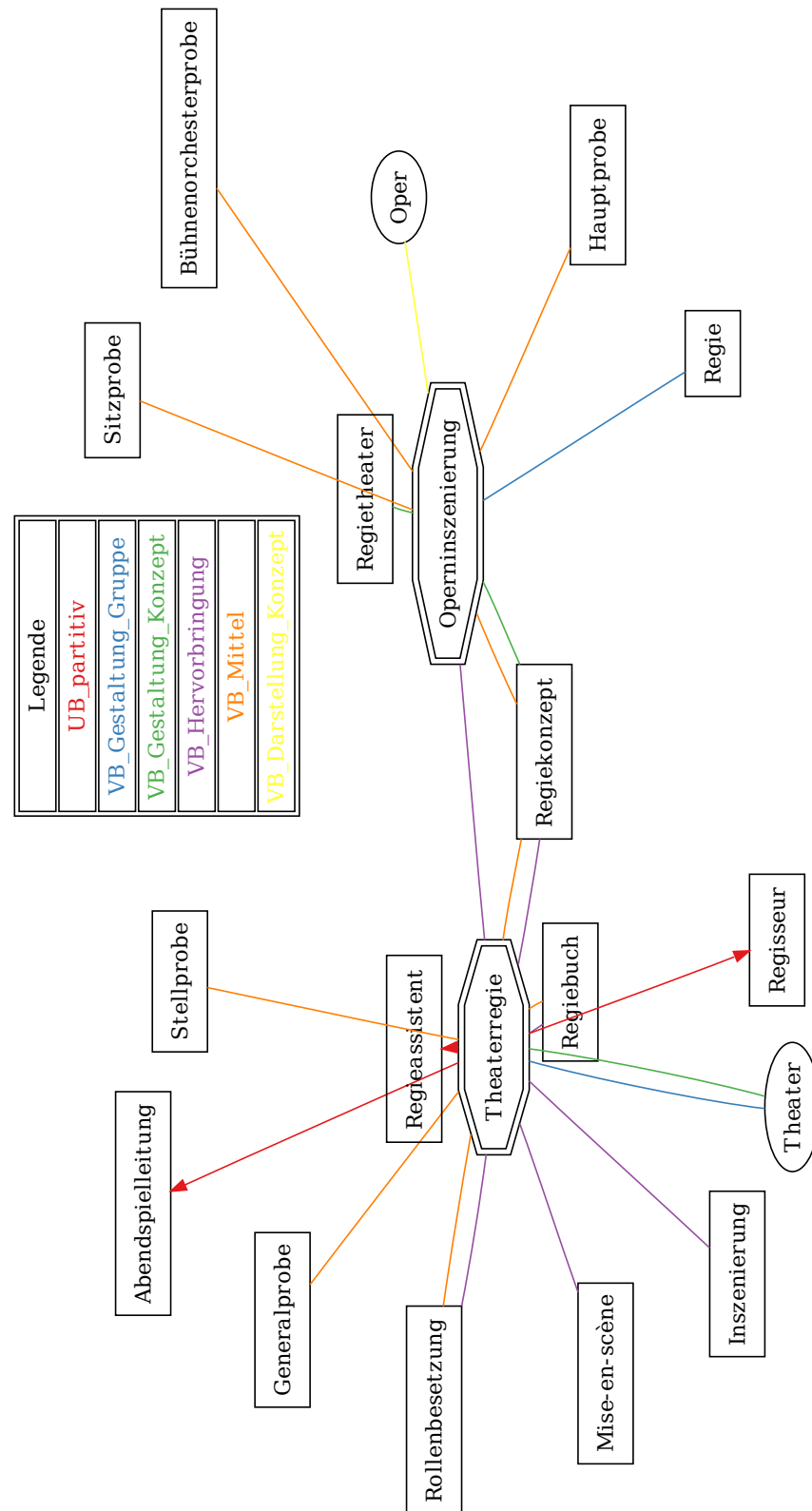


Abbildung 8: Das semantische Umfeld der Begriffe Theaterregie und Operninszenierung (manuell anhand der Skriptausgaben zusammengefügt)

5.5 Entwicklung eines hierarchisch strukturierten themenübergreifenden Relationeninventars

Betrachtet man beide themenspezifischen Inventare zusammen, so zeigen sich in den hierarchischen Relationen große Überschneidungen. In beiden Fällen wird die generische Relation auf der dritten Stufe folgendermaßen verfeinert:

- nach dem Aspekt der Beteiligten (UB_generisch_Beteiligte)
- nach dem Aspekt der Form (UB_generisch_Form)
- nach lokalem Aspekt (UB_generisch_Ort)
- nach zeitlichem Aspekt (UB_generisch_Zeit)

Erst auf der vierten Stufe zeigen sich an manchen Stellen themenspezifische Verfeinerungen – beim Theater etwa nach Stoff und Darstellungsform, bei der Jagd etwa nach eingesetzten sächlichen Mitteln oder eingesetzten Jagdtieren. Es ist also festzuhalten, dass die dritte Stufe, und im Fall der lokalen und zeitlichen Aspekte auch die vierte, themenübergreifende Relationen enthält, während danach themenspezifisch gegliedert wird. Boteram schlägt, wie bereits beschrieben, für solche Fälle vor, in einem themenübergreifenden Vokabular auf den feinen Stufen themenspezifische Ankoppelungen zu erlauben.⁷⁸ Diese werden einerseits dann eingesetzt, wenn ein Relationstyp für einen Gegenstandsbereich ausreichend, für einen anderen jedoch noch zu allgemein ist, und andererseits, wenn ein Relationstyp für mehrere Gegenstandsbereiche zu allgemein ist, sich die Verfeinerungen in den Bereichen jedoch unterscheiden. So kann auch hier verfahren werden: Nach der Relation UB_generisch_Form kann im Theaterbereich das zugeschnittene feine Inventar für jene Thematik angehängt und verwendet werden (Aspekte Aufbau, Stoff und so weiter), im Jagdbereich das entsprechende für diese Thematik (Aspekte Mittel, Tier und so weiter). Ein Aspekt, nämlich der der Methode, erscheint sogar in beiden Inventaren, jedoch in unterschiedlicher, nämlich themenspezifischer Bedeutung.

Wenigere Überschneidungen ergeben sich bei den assoziativen Relationen. Während im Theaterbereich nach Erschaffung, Gestaltung, Darstellung, Betrachtung und sonstiger Beteiligung unterschieden wird, erscheinen im Jagdbereich nur die Beteiligung und die Betrachtung, da weitere Verfeinerungen mangels Anzahl der Relationen in diesem Bedeutungsumfeld nicht nötig geworden sind. Um diese Relationstypen zusammenführen und, wie bei der generischen Relation, eine hierarchische Struktur mit der Möglichkeit der Ankoppelung von themenspezifischen Relationen zu erlauben, müssen diese fünf beziehungsweise zwei Relationstypen nun doch unter einer

⁷⁸Vgl. erneut [Boteram 2008], S. 66–73.

Relation der „Beteiligung“ zusammengeführt werden, was bei der Entwicklung der Theaterinventars ja noch nicht geschehen war. Hierfür ist es nun nötig, den Begriff der Beteiligung zu definieren: Unter ihm soll nun also die aktive Beteiligung an etwas verstanden werden, die aufgrund ihres aktiven Charakters nur durch Menschen erfolgen kann. Hiermit sind einerseits aktive Menschen, andererseits Konzepte der aktiven Beteiligung durch Menschen gemeint. Darunter fallen, angekoppelt im Theaterbereich, die Erschaffung, Gestaltung, Darstellung und Betrachtung, während im Jagdbereich lediglich die Betrachtung erscheint. Die „sonstige Beteiligung“ würde dann ausgedrückt, indem schlicht der allgemeinere Beziehungstyp der „Beteiligung“ verwendet wird. Hierbei sollen auch die Verfeinerungen nach Person, Gruppe, Institution und Konzept, ebenso nach den zugehörigen Instanzen zum themenspezifischen Teil, der angekoppelt wird, gehören, da sie möglicherweise nicht in allen Themenbereichen nötig werden.

Es fällt auf, dass der Beziehungstyp **VB_Mittel** im Jagdbereich verfeinert wird, während dies im Theaterbereich nicht der Fall ist. Dies kann in einem themenübergreifenden Inventar ebenfalls per Ankoppelung gelöst werden. Das Gleiche gilt für **VB_Zeit**. Der Beziehungstyp **VB_Hervorbringung** erscheint in beiden Inventaren, ist jedoch in beiden gleichermaßen schwammig definiert. Dennoch kann er aufgrund seines zweimaligen Vorkommens zunächst in ein allgemeines Inventar übernommen werden, bedarf jedoch in einer realen Anwendung einer genaueren Definition.

Ausschließlich im Theaterinventar erscheint der Relationstyp **VB_Stoff**, ausschließlich im Jagdinventar der Relationstyp **VB_Objekt**. Bei genauerer Betrachtung bezeichnen beide Typen auf einer höheren Abstraktionsebene jedoch das Gleiche: Etwas, mit dem durch die Beteiligten etwas ausgeübt wird, das eigentliche Objekt, das bearbeitet wird. Dies steht im Gegensatz zum Mittel, das eben *für* diese Bearbeitung verwendet wird. Diese Relationstypen ließen sich also zusammenführen, hier ergibt sich jedoch das bereits in den allgemeinen Überlegungen in Abschnitt 4.1 skizzierte Problem, das bei einer solchen Zusammenführung auf eine hohe Abstraktionsebene gewechselt werden muss, die gegebenenfalls nicht mehr allgemeinverständlich ist: Kann etwa der in einem Theaterstück behandelte Stoff als Objekt, an dem etwas ausgeübt wird, bezeichnet werden? Dennoch soll dies hier versucht werden, indem beide Beziehungstypen zu einer Beziehung „Objekt“ zusammengeführt werden, ebenfalls mit der Möglichkeit der themenspezifischen Ankoppelung.

Der Beziehungstyp **VB_Eigenschaft** ist lediglich im Jagdinventar vorhanden. Er ist jedoch bereits allgemeiner Natur, sodass er in ein themenübergreifendes Inventar übernommen werden kann.

Zusammengefasst ergibt sich also folgendes themenübergreifendes Relationeninventar. An allen Stellen sind themenspezifische Ankoppelungen möglich.

- hierarchische Relation
 - generische Relation
 - * nach Aspekt der Form
 - * nach Aspekt der Beteiligten
 - * nach Aspekt des lokalen Bezugs
 - * nach Aspekt des zeitlichen Bezugs
 - partitive Relation
 - Instanzrelation
- Äquivalenzrelation
- assoziative Relation
 - aktive Beteiligung
 - Eigenschaft
 - Hervorbringung
 - Mittel
 - Objekt
 - nicht-hierarchischer zeitlicher Bezug

Festzuhalten ist zunächst, dass dieser allgemeine Teil des Inventars mit 16 Relationstypen keine zu hohe Komplexität aufweist, gleichzeitig jedoch im Gegensatz zu den drei klassischen Relationstypen ein genügendes Maß an Differenziertheit ermöglicht. Das Inventar ist jedoch noch unter der Prämisse der Allgemeinverständlichkeit zu untersuchen, die, wie beschrieben, bei einer Anwendung in der Wikipedia aufgrund ihrer kollaborativen Natur zudem im Vordergrund stehen muss.

Die Verfeinerung der generischen Relation ist sicherlich allgemeinverständlich und mit vier Relationstypen nicht übermäßig komplex. Problematisch sind vor allen Dingen die Relationstypen in der assoziativen Relation. Während die aktive Beteiligung, die Eigenschaft und das Mittel allgemeinverständlich definiert werden können, ist zu bezweifeln, dass die Relationen „Hervorbringung“ und „Objekt“ intuitiv verstanden werden können, ebenso ein „nicht-hierarchischer zeitlicher Bezug“. Hier werden bei einer realen Anwendung des Inventars Umformulierungen und gegebenenfalls Umgestaltungen nötig. Ein endgültiges Inventar kann nur entwickelt werden, wenn neben dem Theater und der Jagd noch weitere Gegenstandsbereiche betrachtet werden, die

weitere Erkenntnisse über die Typen der assoziativen Relation ermöglichen. Dabei ist es gegebenenfalls auch möglich, noch weitere Typen der generischen Relation zu ermitteln, die in den zwei betrachteten Themenbereichen nicht erschienen.

Die vor Beginn der Relationierungsarbeit aufgeworfene Frage, ob es überhaupt möglich ist, ein themenübergreifendes Relationeninventar zu entwickeln, kann jedoch grundsätzlich positiv beantwortet werden, sofern die Möglichkeit der Ankoppelung themenspezifischer Typen gegeben ist. Problematisch ist vor allen Dingen die assoziative Relation, hier ist darauf zu achten, dass keine zu abstrakten Relationstypen formuliert werden.

Betreffend letztere Relation fällt zudem im Vergleich mit den von Boteram und Michel entwickelten Relationeninventaren auf, dass die Zahl der assoziativen Relationstypen im oben dargestellten Inventar relativ klein ist. Dies ist der Tatsache geschuldet, dass in den zwei bearbeiteten Gegenstandsbereichen nur die bereits vorhandenen Relationen typisiert wurden. Assoziative Relationen anderer Ausprägungen waren schlicht nicht vorhanden, obgleich es beim Aufbau der beiden Kategorienteilsysteme sicherlich möglich gewesen wäre, zwischen den vorhandenen Begriffen weitere Relationen assoziativer Natur herausarbeiten. Bei der zukünftigen Aufnahme neuer Relationen in die zwei Vokabulare kann es jedoch passieren, dass assoziative Relationen erscheinen, die von den jetzt formulierten Relationstypen in keinerlei Hinsicht abgedeckt werden. Dann kann das themenübergreifende Inventar um weitere, völlig neue assoziative Relationstypen auf der zweiten Hierarchiestufe ergänzt werden.

6 Schluss

Die vorliegende Arbeit beschäftigte sich mit der Natur des Kategoriensystem der Wikipedia sowie exemplarisch anhand zweier Gegenstandsbereiche mit den im Kategoriensystem der deutschsprachigen Wikipedia vorkommenden semantischen Relationen zwischen Allgemeinbegriffen. Es konnte gezeigt werden, dass in den zwei untersuchten Teilsystemen eine hohe Zahl von Relationen verschiedenster Typen vorkommen, deren semantischer Gehalt jedoch bislang nicht näher beschrieben wurde.

Bei den Untersuchungen wurde festgestellt, dass das Kategoriensystem, zumindest in bestimmten Teilen, zu denen die zwei untersuchten gehören, hoch entwickelt ist und bei einer Betrachtung der Kategorien- und Artikelnamen als Begriffe ein großes und umfangreich relationiertes Begriffssystem darstellt.

Würden die Relationen nicht wie bislang mittels eines einzigen Relationstyps dargestellt, sondern ihr semantischer Gehalt formal anhand eines differenzierten Relationeninventars festgehalten, so bedeutete dies eine umfangreiche Erweiterung der Navigations- und Suchmöglichkeiten im Kategoriensystem und damit in der Wikipedia. Diese Möglichkeiten versuchte diese Arbeit darzulegen, indem differenzierte Relationeninventare zur gemeinschaftlichen Anwendung im Kategoriensystem geschaffen wurden – zunächst zwei gegenstandsspezifische und daraus in einer ersten Annäherung ein Vorschlag für ein themenübergreifendes –, sämtliche in den beiden Bereichen vorhandenen Relationen mit den neuen Inventaren typisiert, und die neuen Strukturen exemplarisch visualisiert wurden.

Ergebnis der Arbeit ist ein themenübergreifendes, hierarchisch strukturiertes Relationeninventar. Sofern die technischen Voraussetzungen zur realen Anwendung in der Wikipedia geschaffen werden, können weitere Gegenstandsbereiche anhand des neuen Inventars relationiert werden, wobei das Inventar gegebenenfalls noch erweitert werden muss, was leicht möglich ist. Sobald große Teile des Kategoriensystems differenziert relationiert sind – in den bereits gut entwickelten Teilen bedeutet dies in der Regel nur das Typisieren der vorhandenen Relationen, nicht notwendigerweise das Schaffen neuer oder das Löschen falscher –, können die skizzierten neuen Möglichkeiten in Navigation und Suche umgesetzt werden. Sind bislang die Hyperlinks in den Artikeltexten die wohl prominenteste Navigationsmöglichkeit in der Wikipedia, so können mit den neuen Strukturen im Kategoriensystem neue Möglichkeiten der systematischen Exploration der Wikipedia gegeben werden. Das Kategoriensystem kann sich so als leistungsfähiges Instrument zur Exploration jeglicher Themenbereiche etablieren.

Einschränkend ist jedoch zu sagen, dass einige Teile des Kategoriensystems der deutschsprachigen Wikipedia, insbesondere einige, die außerhalb der sogenannten Fachbereiche liegen, noch starken Überarbeitungen bedürfen, die über das Typisieren

der vorhandenen Relationen hinausgehen. An diesen Stellen ist höherer Aufwand nötig, dennoch besteht an jeder Stelle des Kategoriensystems die Möglichkeit zur Transformation in ein hoch entwickeltes und differenziert relationiertes Vokabular, das den oben genannten Navigations- und Suchmöglichkeiten gerecht wird.

Literatur

Auf sämtliche Internetquellen wurde zum letzten Mal am 24. August 2010 zugegriffen.

- [Aitchison 2000] Aitchison, Jean; Alan Gilchrist, David Bawden: Thesaurus construction and use. A practical manual. 4th ed. London 2000.
- [Bertram 2005] Bertram, Jutta: Einführung in die inhaltliche Erschließung. Grundlagen, Methoden, Instrumente. Würzburg 2005. (Content and communication. Terminology, language resources and semantic interoperability; Bd. 2.)
- [Boteram 2008] Boteram, Felix: Semantische Relationen in Dokumentations-sprachen. Vom Thesaurus zum semantischen Netz. Köln 2008. (Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft; Bd. 54.)
<http://www.fbi.fh-koeln.de/institut/papers/kabi/volltexte/band054.pdf>
- [Catgraph 2009] Catgraph. 2009.
<http://toolserver.org/~dapete/catgraph/>
- [DBpedia 2010] DBpedia. Leipzig [u.a.] 2010.
<http://dbpedia.org/>
- [Faviki 2010] Faviki. Belgrad 2010.
<http://www.faviki.com/>
- [Frost 2006] Frost, Ingo: Zivilgesellschaftliches Engagement in virtuellen Gemeinschaften? Eine systemwissenschaftliche Analyse des deutschsprachigen Wikipedia-Projektes. München 2006. Zugl.: Osnabrück, Univ., Diplomarbeit, 2006. Auch online unter:
http://www-lehre.informatik.uni-osnabrueck.de/~ifrost/offiziell/frost2006_wikipedia.pdf
- [Graphviz 2010] Graphviz. Graph Visualization Software. 2010.
<http://www.graphviz.org/>
- [Hubrich 2009] Hubrich, Jessica: Vom Stringmatching zur Begriffsexploration. Das Potential integrierter begrifflicher Interoperabilität. In: Tagungsband der 12. Tagung der Deutschen Sektion der ISKO 2009 in Bonn. Würzburg [noch nicht erschienen]. Auch online unter:

- http://linux2.fbi.fh-koeln.de/crisscross/publikationen/Hubrich_IntegrierteBegrifflicheInteroperabilitaet.pdf
- [MACS 2010] The MACS Project. Enabling large-scale multilingual access to subjects.
<http://macs.hoppie.nl/>
- [MediaWiki 2010] Database layout. In: MediaWiki technical manual. 2010.
http://www.mediawiki.org/wiki/Manual:Database_layout
- [Meibauer 2002] Meibauer, Jörg; Ulrike Demske, Jochen Geilfuß-Wolfgang, Jürgen Pafel, Karl Heinz Ramers, Monika Rothweiler, Markus Steinbach: Einführung in die germanistische Linguistik. Stuttgart 2002.
- [Michel 1997] Michel, Dee: Taxonomy of subject relationships. Appendix B zu: Association for Library Collections & Technical Services, Subcommittee on Subject Relationships / Reference Structure (Hrsg.): Final Report to the ALCTS/CSS Subject Analysis Committee. 1997.
<http://www.ala.org/ala/mgrps/divs/alcts/mgrps/ccs/cmtes/sac/inact/subjrel/msrscu2.pdf>
- [Newham 2005] Newham, Cameron; Bill Rosenblatt: Einführung in die bash-Shell. Dt. Übers. von Dinu C. Gherman. Sebastopol [u.a.] 2005. (Learning the bash shell; dt.)
- [NISO 2005] National Information Standards Organization (Hrsg.): Guidelines for the construction, format, and management of monolingual Controlled Vocabularies. ANSI/NISO Z39.19-2005.
- [Reagle 2004] Reagle, Joseph M., Jr.: A case of mutual aid. Wikipedia, politeness and perspective taking. 2004.
<http://reagle.org/joseph/2004/agree/wikip-agree.html>
- [Suchanek 2007] Suchanek, Fabian M.; Gjergji Kasneci, Gerhard Weikum: YAGO. A core of semantic knowledge unifying WordNet and Wikipedia. In: WWW2007. Proceedings of the 16th international conference on World Wide Web. New York 2007. S. 697–706. Auch online unter:
<http://www2007.org/papers/paper391.pdf>

- [Voß 2004a] Voß, Jakob: Categories considered harmful. In: Wikipedia-l. 2004.
<http://lists.wikimedia.org/pipermail/wikipedia-l/2004-June/015857.html>
- [Voß 2004b] Benutzer:JakobVoss / Categories considered harmful. In: Wikipedia. Die freie Enzyklopädie. 2004.
http://de.wikipedia.org/w/index.php?title=Benutzer:JakobVoss/Categories_considered_harmful&oldid=2713920
- [Voß 2005] Voß, Jakob: Informetrische Untersuchungen an der Online-Enzyklopädie Wikipedia. Berlin, Humboldt-Univ., Magisterarbeit. Berlin 2005. Auch online unter:
<http://jakobvoss.de/magisterarbeit/MagisterarbeitJakobVoss.pdf>
- [Voß 2006] Voß, Jakob: Collaborative thesaurus tagging the Wikipedia way. 2006. (Wikimetrics research papers, Bd. 1, Nr. 1.)
<http://arxiv.org/pdf/cs/0604036v2>
- [Weller 2006] Weller, Katrin: Kooperativer Ontologieaufbau. In: Ockenfeld, Marlies (Hrsg.): Content. 28. Online-Tagung der DGI, 58. Jahrestagung der DGI. Proceedings. Frankfurt (Main) 2006. S. 227–234. Auch online unter:
http://wwwalt.phil-fak.uni-duesseldorf.de/infowiss/admin/public_dateien/files/35/1166536740katrinwell.pdf
- [Wersig 1978] Wersig, Gernot: Thesaurus-Leitfaden. Eine Einführung in das Thesaurus-Prinzip in Theorie und Praxis. München [u.a.] 1978 (DGD-Schriftenreihe; Bd. 8.)
- [Wikimedia 2010] Wikimedia Downloads. 2010.
<http://dumps.wikimedia.org/>
- [Wikipedia 2010a] Wikipedia. In: Wikipedia. Die freie Enzyklopädie. 2010.
<http://de.wikipedia.org/w/index.php?title=Wikipedia&oldid=78116384>
- [Wikipedia 2010b] Wikipedia: Sprachen. In: Wikipedia. Die freie Enzyklopädie. 2010.
<http://de.wikipedia.org/w/index.php?title=Wikipedia:Sprachen&oldid=78155988>

- [Wikipedia 2010c] Ontologie (Informatik). In: Wikipedia. Die freie Enzyklopädie. 2010.
[http://de.wikipedia.org/w/index.php?title=Ontologie_\(Informatik\)&oldid=77133128](http://de.wikipedia.org/w/index.php?title=Ontologie_(Informatik)&oldid=77133128)
- [Wikipedia 2010d] Wikipedia: Namenskonventionen. In: Wikipedia. Die freie Enzyklopädie. 2010.
<http://de.wikipedia.org/w/index.php?title=Wikipedia:Namenskonventionen&oldid=78097054>
- [Wikipedia 2010e] Wikipedia: Kategorien. In: Wikipedia. Die freie Enzyklopädie. 2010.
<http://de.wikipedia.org/w/index.php?title=Wikipedia:Kategorien&oldid=77188791>
- [Wikipedia 2010f] Wikipedia: WikiProjekt Kategorien / Fachbereiche. In: Wikipedia. Die freie Enzyklopädie. 2010.
http://de.wikipedia.org/w/index.php?title=Wikipedia:WikiProjekt_Kategorien/Fachbereiche&oldid=78042743
- [Wikipedia 2010g] Wikipedia: Autorenportal. In: Wikipedia. Die freie Enzyklopädie. 2010.
<http://de.wikipedia.org/w/index.php?title=Wikipedia:Autorenportal&oldid=78097665>
- [Wikipedia 2010h] Wikipedia: WikiProjekt Kategorien. In: Wikipedia. Die freie Enzyklopädie. 2010.
http://de.wikipedia.org/w/index.php?title=Wikipedia:WikiProjekt_Kategorien&oldid=78141791
- [Wikipedia 2010i] Wikipedia Diskussion: Kategorien. In: Wikipedia. Die freie Enzyklopädie. 2010.
http://de.wikipedia.org/w/index.php?title=Wikipedia_Diskussion:Kategorien&oldid=78156634
- [Wikipedia 2010k] Wikipedia: WikiProjekt Kategorien / Einordnung von Kategorien. In: Wikipedia. Die freie Enzyklopädie. 2010.
http://de.wikipedia.org/w/index.php?title=Wikipedia:WikiProjekt_Kategorien/Einordnung_von_Kategorien&oldid=78089447
- [WordNet 2010] WordNet. A lexical database for English. Princeton 2010.
<http://wordnet.princeton.edu/>

Anhang

A Relationeninventar für den Gegenstandsbereich Theater

In der folgenden Tabelle wird das Relationeninventar für den Gegenstandsbereich Theater dargestellt. Die linken Spalten der jeweils ersten Zeile benennen den jeweiligen Relationstypen nach Hierarchiestufen getrennt. Die rechte Spalte gibt seine Bezeichnung im Vokabular an. Bleibt die rechte Spalte leer, so wurde der betreffende Relationstyp lediglich zur Hierarchiebildung eingefügt und kommt im Vokabular nicht vor. In der jeweils zweiten Zeile steht eine knappe Erläuterung des Relationstypen.

Stufe				
1	2	3	4	Bezeichnung in Theater-bearbeitet.txt
TT				TT
	Topterm			
UB				
	allgemeine Unterbegriffsrelation			
UB	generisch			
	allgemeine generische Unterbegriffsrelation			
UB	generisch	Beteiligte		UB_generisch_Beteiligte
	nach Aspekt der Beteiligten im Allgemeinen			
UB	generisch	Beteiligte	Darstellende	UB_generisch_Beteiligte_Darstellende
	nach Aspekt der Darstellenden			
UB	generisch	Beteiligte	Publikum	UB_generisch_Beteiligte_Publikum
	nach Aspekt des Publikums			
UB	generisch	Form		UB_generisch_Form
	nach Aspekt der Form im Allgemeinen			
UB	generisch	Form	Aufbau	UB_generisch_Form_Aufbau
	nach Aspekt des Aufbaus (Zahl und äußere Form der Akte, Szenen usw.)			
UB	generisch	Form	Darstellung	UB_generisch_Form_Darstellung
	nach Aspekt der Darstellungsform			
UB	generisch	Form	Methode	UB_generisch_Form_Methode
	nach Aspekt der grundlegenden Methode (Improvisation, Drama mit Gesangsteilen usw.)			
UB	generisch	Form	Ort	UB_generisch_Form_Ort
	nach Aspekt des Darstellungsortes			
UB	generisch	Form	Sparte	UB_generisch_Form_Sparte
	nach Aspekt der Sparte (betrifft nur die vier klassischen Theatersparten)			
UB	generisch	Form	Stoff	UB_generisch_Form_Stoff
	nach Aspekt des behandelten Stoffs (inhaltlich)			
UB	generisch	Ort		
	nach Aspekt eines lokalen Bezugs im Allgemeinen			

UB	generisch	Ort	geographisch	UB_generisch_Ort_geographisch
nach Aspekt des Bezugs zu einem geographischen Ort				
UB	generisch	Zeit		
nach Aspekt eines zeitlichen Bezugs im Allgemeinen				
UB	generisch	Zeit	Anlass	UB_generisch_Zeit_Anlass
nach Aspekt des Bezugs zu einem Anlass				
UB	generisch	Zeit	historisch	UB_generisch_Zeit_historisch
nach Aspekt des Bezugs zu einer historischen Zeitspanne				
UB	partitiv			UB_partitiv
partitive Relation				
UB	Instanz			UB_Instanz
Instanzrelation				
VB				
assoziative Relation (verwandter Begriff)				
VB	Erschaffung			
beteiligt an kreativem Prozess der Erschaffung				
VB	Erschaffung	Person		VB_Erschaffung_Person
eine Person, die kreativ erschafft (Dramatiker, Librettist usw.)				
VB	Erschaffung	Person	Instanz	VB_Erschaffung_Person_Instanz
eine Instanz einer Person, die kreativ erschafft				
VB	Erschaffung	Gruppe		
eine Gruppe, die kreativ erschafft				
VB	Erschaffung	Gruppe	Instanz	VB_Erschaffung_Gruppe_Instanz
eine Instanz einer Gruppe, die kreativ erschafft				
VB	Erschaffung	Konzept		VB_Erschaffung_Konzept
ein Konzept in der kreativen Erschaffung (Choreografie, Kreatives Schreiben usw.)				
VB	Gestaltung			
beteiligt an der Gestaltung von etwas kreativ Erschaffenem				
VB	Gestaltung	Person		VB_Gestaltung_Person
eine Person, die gestaltet im o. g. Sinne (Regisseur, Beleuchter usw.)				
VB	Gestaltung	Person	Instanz	VB_Gestaltung_Person_Instanz
eine Instanz einer Person, die gestaltet im o. g. Sinne				
VB	Gestaltung	Gruppe		VB_Gestaltung_Gruppe
eine Gruppe, die gestaltet im o. g. Sinne				
VB	Gestaltung	Gruppe	Instanz	VB_Gestaltung_Gruppe_Instanz
eine Instanz einer Gruppe, die gestaltet im o. g. Sinne				
VB	Gestaltung	Konzept		VB_Gestaltung_Konzept
ein Konzept der Gestaltung im o. g. Sinne (Bühnentechnik als Konzept usw.)				
VB	Gestaltung	Ort		VB_Gestaltung_Ort
ein Ort zur Gestaltung im o. g. Sinne				

VB	Gestaltung	Ort	Instanz	VB_Gestaltung_Ort_Instanz
	eine Instanz eines Ortes zur Gestaltung im o. g. Sinne			
VB	Darstellung			
	beteiligt an der Darstellung von etwas Erschaffenem			
VB	Darstellung	Gruppe		VB_Darstellung_Gruppe
	eine Gruppe, die darstellt			
VB	Darstellung	Gruppe	Instanz	VB_Darstellung_Gruppe_Instanz
	eine Instanz einer Gruppe, die darstellt			
VB	Darstellung	Institution		
	eine Institution, die darstellt bzw. in der dargestellt wird			
VB	Darstellung	Institution	Instanz	VB_Darstellung_Institution_Instanz
	eine Instanz einer Institution, die darstellt bzw. in der dargestellt wird			
VB	Darstellung	Konzept		VB_Darstellung_Konzept
	ein Konzept der Darstellung (Akt, Abgang usw.)			
VB	Darstellung	Ort		VB_Darstellung_Ort
	ein Ort, an dem dargestellt wird			
VB	Darstellung	Ort	Instanz	VB_Darstellung_Ort_Instanz
	eine Instanz eines Ortes, an dem dargestellt wird			
VB	Darstellung	Person		VB_Darstellung_Person
	eine Person, die darstellt (Schauspieler, Tänzer usw.)			
VB	Darstellung	Person	Instanz	VB_Darstellung_Person_Instanz
	eine Instanz einer Person, die darstellt			
VB	Betrachtung			
	beteiligt an einer Betrachtung von außen (Theaterwissenschaft, -kritik usw.)			
VB	Betrachtung	Gegenstand		VB_Betrachtung_Gegenstand
	ein Gegenstand, der betrachtet wird im o. g. Sinne			
VB	Betrachtung	Institution		VB_Betrachtung_Institution
	eine Institution, die etwas betrachtet im o. g. Sinne			
VB	Betrachtung	Institution	Instanz	VB_Betrachtung_Institution_Instanz
	eine Instanz einer Institution, die etwas betrachtet im o. g. Sinne			
VB	Betrachtung	Konzept		VB_Betrachtung_Konzept
	ein Konzept in der Betrachtung im o. g. Sinne			
VB	Hervorbringung			VB_Hervorbringung
	eine weit gefasste Produzent-Produkt-Beziehung			
VB	Beteiligung			
	sonstige Beteiligung außer den oben genannten Formen			
VB	Beteiligung	Person		VB_Beteiligung_Person
	eine Person, die sonst beteiligt ist			
VB	Beteiligung	Person	Instanz	VB_Beteiligung_Person_Instanz
	eine Instanz einer Person, die sonst beteiligt ist			

VB	Beteiligung	Institution		VB_Beteiligung_Institution
	eine Institution, die sonst beteiligt ist			
VB	Beteiligung	Institution	Instanz	VB_Beteiligung_Institution_Instanz
	eine Instanz einer Institution, die sonst beteiligt ist			
VB	Beteiligung	Konzept		VB_Beteiligung_Konzept
	ein Konzept in einer sonstigen Beteiligung			
VB	Mittel			VB_Mittel
	ein Mittel, das benutzt bzw. eingesetzt wird			
VB	Stoff			VB_Stoff
	Stoff (Inhaltliches), der behandelt wird			
VB	Stoff	Konzept		VB_Stoff_Konzept
	ein Konzept in der Behandlung von Stoff im o. g. Sinne			
VB	Zeit	hist. Ereignis	Instanz	VB_Zeit_historischesEreignis_Instanz
	ein konkretes historisches Ereignis, das in Beziehung zu etwas anderem steht			

Die folgende Tabelle enthält Angaben zur Häufigkeit der Relationstypen im vorliegenden Vokabular zum Gegenstandsbereich Theater.

Anzahl	Relationstyp
12102	UB_Instanz
2332	VB_Erschaffung_Person_Instanz
241	--
145	VB_Betrachtung_Gegenstand
104	VB_Mittel
104	UB_generisch_Ort_geographisch
103	UB_partitiv
78	UB_generisch_Form_Stoff
77	UB_generisch_Form_Aufbau
56	VB_Darstellung_Person_Instanz
48	UB_generisch_Zeit_historisch
41	VB_Darstellung_Ort_Instanz
41	UB_generisch_Form
39	VB_Gestaltung_Person
37	VB_Darstellung_Konzept
36	VB_Gestaltung_Person_Instanz
36	UB_generisch_Form_Methode
31	VB_Betrachtung_Institution_Instanz
30	VB_Stoff
29	VB_Beteiligung_Institution_Instanz
25	VB_Hervorbringung
23	VB_Darstellung_Ort
21	VB_Darstellung_Gruppe_Instanz
21	UB_generisch_Form_Darstellung
19	UB_generisch_Form_Sparte
16	VB_Darstellung_Person
15	VB_Gestaltung_Ort
14	UB_generisch_Form_Ort
13	UB_generisch_Beteiligte_Publikum
12	VB_Erschaffung_Person
11	VB_Darstellung_Gruppe
11	UB_generisch_Beteiligte_Darstellende
8	VB_Gestaltung_Konzept
7	VB_Zeit_historischesEreignis_Instanz
7	VB_Stoff_Konzept
7	VB_Gestaltung_Ort_Instanz
6	UB_generisch_Zeit_Anlass
5	VB_Gestaltung_Gruppe
5	VB_Betrachtung_Konzept
4	VB_Erschaffung_Konzept
4	VB_Beteiligung_Institution
3	VB_Gestaltung_Gruppe_Instanz
3	VB_Darstellung_Institution_Instanz

2	VB_Beteiligung_Person
2	VB_Beteiligung_Konzept
2	UB_generisch_Beteiligte
1	VB_Erschaffung_Gruppe_Instance
1	VB_Betrachtung_Institution
1	VB_Beteiligung_Person_Instance
1	TT
15980	Summe

B Relationeninventar für den Gegenstandsbereich Jagd

In der folgenden Tabelle wird das Relationeninventar zum Gegenstandsbereich Jagd dargestellt. Es gelten die gleichen Anmerkungen wie zu ihrer Entsprechung in Anhang A auf S. 73.

Stufe				
1	2	3	4	Bezeichnung in Jagd-bearbeitet.txt
TT	Topterm			TT
UB	allgemeine Unterbegriffsrelation			
UB	generisch			
	allgemeine generische Unterbegriffsrelation			
UB	generisch	Beteiligte		UB_generisch_Beteiligte
	nach Aspekt der Beteiligten			
UB	generisch	Form		UB_generisch_Form
	nach Aspekt der Form im Allgemeinen			
UB	generisch	Form	Eigenschaft	UB_generisch_Form_Eigenschaft
	nach Aspekt einer bestimmten Eigenschaft			
UB	generisch	Form	Funktion	UB_generisch_Form_Funktion
	nach Aspekt der Funktion			
UB	generisch	Form	Klasse	UB_generisch_Form_Funktion
	nach Aspekt der Tierklasse (Säugetiere, Vögel usw.)			
UB	generisch	Form	Methode	UB_generisch_Form_Methode
	nach Aspekt der Methode			
UB	generisch	Form	Tier	UB_generisch_Form_Tier
	nach Aspekt der Tierart bzw. -gattung			
UB	generisch	Ort		
	nach Aspekt eines lokalen Bezugs im Allgemeinen			
UB	generisch	Ort	geographisch	UB_generisch_Ort_geographisch
	nach Aspekt des Bezugs zu einem geographischen Ort			
UB	generisch	Zeit		
	nach Aspekt eines zeitlichen Bezugs im Allgemeinen			
UB	generisch	Zeit	historisch	UB_generisch_Zeit_historisch
	nach Aspekt des Bezugs zu einer historischen Zeitspanne			
UB	partitiv			UB_partitiv
	partitive Relation			
UB	Instanz			UB_Instanz
	Instanzrelation			
VB	assoziative Relation (verwandter Begriff)			

VB	Betrachtung			VB_Betrachtung
	beteiligt an einer Betrachtung von außen			
VB	Betrachtung	Person		
	eine Person, die etwas betrachtet im o. g. Sinne			
VB	Betrachtung	Person	Instanz	VB_Betrachtung_Person_Instanz
	eine Instanz einer Person, die etwas betrachtet im o. g. Sinne			
VB	Betrachtung	Institution		VB_Betrachtung_Institution
	eine Institution, die etwas betrachtet im o. g. Sinne			
VB	Betrachtung	Institution	Instanz	VB_Betrachtung_Institution_Instanz
	eine Instanz einer Institution, die etwas betrachtet im o. g. Sinne			
VB	Betrachtung	Konzept		VB_Betrachtung_Konzept
	ein Konzept in der Betrachtung im o. g. Sinne			
VB	Betrachtung	Ort		VB_Betrachtung_Ort
	ein Ort zur Betrachtung im o. g. Sinne			
VB	Betrachtung	Ort	Instanz	VB_Betrachtung_Ort_Instanz
	eine Instanz eines Ortes zur Betrachtung im o. g. Sinne			
VB	Hervorbringung			VB_Hervorbringung
	eine weit gefasste Produzent-Produkt-Beziehung			
VB	Beteiligung			VB_Beteiligung
	sonstige Beteiligung			
VB	Beteiligung	Person		VB_Beteiligung_Person
	eine Person, die sonst beteiligt ist			
VB	Beteiligung	Person	Instanz	VB_Beteiligung_Person_Instanz
	eine Instanz einer Person, die sonst beteiligt ist			
VB	Beteiligung	Gruppe		VB_Beteiligung_Gruppe
	eine Gruppe, die sonst beteiligt ist			
VB	Beteiligung	Institution		VB_Beteiligung_Institution
	eine Institution, die sonst beteiligt ist			
VB	Beteiligung	Institution	Instanz	VB_Beteiligung_Institution_Instanz
	eine Instanz einer Institution, die sonst beteiligt ist			
VB	Beteiligung	Konzept		VB_Beteiligung_Konzept
	ein Konzept in einer sonstigen Beteiligung			
VB	Beteiligung	Ort		VB_Beteiligung_Ort
	ein Ort zur sonstigen Beteiligung			
VB	Eigenschaft			VB_Eigenschaft
	eine Eigenschaft			
VB	Mittel			
	ein Mittel, das benutzt bzw. eingesetzt wird			
VB	Mittel	Sache		VB_Mittel_Sache
	eine Sache, die benutzt bzw. eingesetzt wird			

VB	Mittel	Tier		VB_Mittel_Tier
	ein Tier, das benutzt bzw. eingesetzt wird			
VB	Mittel	Vorgang		VB_Mittel_Vorgang
	ein Vorgang, um etwas zu erreichen			
VB	Objekt			VB_Objekt
	ein Objekt, an dem etwas getan wird (bejagtes Tier usw.)			
VB	Objekt	Instanz		VB_Objekt_Instanz
	eine Instanz eines Objektes, an dem etwas getan wird			
VB	Zeit			VB_Zeit
	nicht-hierarchische Beziehung mit zeitlichem Bezug			
VB	Zeit	Tradition		VB_Zeit_Tradition
	eine Tradition (jagdliches Brauchtum usw.)			
VB	Zeit	hist. Betrachtung		VB_Zeit_historischeBetrachtung
	eine historische Betrachtung (bildliche Darstellungen von Jagd usw.)			
VB	Zeit	hist. Betrachtung	Instanz	VB_Zeit_historischeBetrachtung_Instanz
	eine Instanz einer historischen Betrachtung			
VB	Zeit	hist. Ereignis	Instanz	VB_Zeit_historischesEreignis_Instanz
	ein konkretes historisches Ereignis, das in Beziehung zu etwas anderem steht			

Die folgende Tabelle enthält Angaben zur Häufigkeit der Relationstypen im vorliegenden Vokabular zum Gegenstandsbereich Jagd.

Anzahl	Relationstyp
329	UB_Instanz
154	UB_partitiv
120	VB_Beteiligung_Person_Instanz
89	UB_generisch_Form
31	UB_generisch_Form_Methode
28	UB_generisch_Form_Tier
27	VB_Mittel_Sache
22	--
14	VB_Beteiligung_Institution_Instanz
13	VB_Mittel_Vorgang
13	UB_generisch_Form_Mittel
11	VB_Zeit_Tradition
10	VB_Zeit_historischeBetrachtung_Instanz
9	VB_Beteiligung_Person
8	VB_Hervorbringung
8	VB_Betrachtung_Person_Instanz
8	UB_generisch_Beteiligte
7	VB_Objekt
7	UB_generisch_Form_Funktion
6	VB_Betrachtung_Institution_Instanz
5	UB_generisch_Zeit_historisch
4	VB_Mittel_Tier
4	VB_Beteiligung_Institution
3	VB_Zeit
3	VB_Eigenschaft
2	VB_Betrachtung_Ort_Instanz
2	VB_Beteiligung_Ort
2	VB_Beteiligung_Gruppe
2	UB_generisch_Ort_geographisch
2	UB_generisch_Form_Klasse
1	VB_Zeit_historischesEreignis_Instanz
1	VB_Zeit_historischeBetrachtung
1	VB_Objekt_Instanz
1	VB_Betrachtung_Ort
1	VB_Betrachtung_Konzept
1	VB_Betrachtung_Institution
1	VB_Betrachtung
1	VB_Beteiligung_Konzept
1	VB_Beteiligung
1	UB_generisch_Form_Eigenschaft
1	TT
954	Summe

C Visualisierungen mit Graphviz

C.1 Quelltexte zu den Abbildungen 6 und 7

Die abgedruckten Quelltexte wurden von den Skripten `visualisierung-or` beziehungsweise `visualisierung-and` erzeugt und sind, zusammen mit sämtlichen verarbeiteten Visualisierungsbeispielen, auch auf der dieser Arbeit beigelegten CD-ROM im Verzeichnis `visualisierung/` enthalten.

Semantisches Umfeld des Begriffs Theaterregie (Theaterregie-OR.gv, zu Abbildung 6 auf S. 59)

```
# Vokabular aus Datei theater/Theater-bearbeitet.txt
# Begriff: Theaterregie
# Relationstypen: UB_partitiv VB_Gestaltung_Gruppe ↓
VB_Gestaltung_Konzept VB_Hervorbringung VB_Mittel
# OR-Verknuepfung

digraph Theaterregie {

Theaterregie [label="Theaterregie", shape=doubleoctagon]

Operninszenierung [label="Operninszenierung" shape=ellipse]
Theater [label="Theater" shape=ellipse]
a_Abendspielleitung [label="Abendspielleitung" shape=box]
a_Generalprobe [label="Generalprobe" shape=box]
a_Inszenierung [label="Inszenierung" shape=box]
a_Mise_en_scène [label="Mise-en-scène" shape=box]
a_Regieassistent [label="Regieassistent" shape=box]
a_Regiebuch [label="Regiebuch" shape=box]
a_Regiekonzept [label="Regiekonzept" shape=box]
a_Regisseur [label="Regisseur" shape=box]
a_Rollenbesetzung [label="Rollenbesetzung" shape=box]
a_Stellprobe [label="Stellprobe" shape=box]
Theaterregie -> a_Abendspielleitung [color="/set19/1"]
Theaterregie -> a_Regieassistent [color="/set19/1"]
Theaterregie -> a_Regisseur [color="/set19/1"]
Theater -> Theaterregie [color="/set19/2" dir=none]
Theater -> Theaterregie [color="/set19/3" dir=none]
Theaterregie -> Operninszenierung [color="/set19/4" dir=none]
Theaterregie -> a_Inszenierung [color="/set19/4" dir=none]
Theaterregie -> a_Mise_en_scène [color="/set19/4" dir=none]
Theaterregie -> a_Regiebuch [color="/set19/4" dir=none]
Theaterregie -> a_Regiekonzept [color="/set19/4" dir=none]
Theaterregie -> a_Rollenbesetzung [color="/set19/4" dir=none]
Theaterregie -> a_Generalprobe [color="/set19/5" dir=none]
```

```

Theaterregie -> a_Regiebuch [color="/set19/5" dir=none]
Theaterregie -> a_Regiekonzept [color="/set19/5" dir=none]
Theaterregie -> a_Rollenbesetzung [color="/set19/5" dir=none]
Theaterregie -> a_Stellprobe [color="/set19/5" dir=none]

legende [label=<
<table cellpadding="4">
<tr><td>Legende</td></tr>
<tr><td><font color="/set19/1">UB_partitiv</font></td></tr>
<tr><td><font color="/set19/2">VB_Gestaltung_Gruppe</font></td></tr>
<tr><td><font color="/set19/3">VB_Gestaltung_Konzept</font></td></tr>
<tr><td><font color="/set19/4">VB_Hervorbringung</font></td></tr>
<tr><td><font color="/set19/5">VB_Mittel</font></td></tr>
</table>
>, shape=plaintext]

}

```

Unterbegriffe des Begriffs Theater nach Stoff und historischem Bezug (Dateiname Theater-Stoff-Zeit-AND.gv, zu Abbildung 7 auf S. 60)

```

# Vokabular aus Datei theater/Theater-bearbeitet.txt
# Begriff: Theater
# Relationstypen: UB_generisch_Form_Stoff ↓
UB_generisch_Zeit_historisch
# AND-Verknuepfung

digraph Theater {

Theater [label="Theater", shape=doubleoctagon]

a_Absurdes_Theater [label="Absurdes Theater" shape=box]
a_Arbeiter__und_Studententheater [label="Arbeiter- und ↓
Studententheater" shape=box]
a_Besserungsstück [label="Besserungsstück" shape=box]
a_Fabula_praetexta [label="Fabula praetexta" shape=box]
a_Fabula_togata [label="Fabula togata" shape=box]
a_Lehrstück__Theater_ [label="Lehrstück (Theater)" shape=box]
a_Moralitäten [label="Moralitäten" shape=box]
a_Musical_Comedy [label="Musical Comedy" shape=box]
a_Musical_Play [label="Musical Play" shape=box]
a_Neidhartspiel [label="Neidhartspiel" shape=box]
a_Palliata [label="Palliata" shape=box]
a_Thingspiel [label="Thingspiel" shape=box]
Theater -> a_Absurdes_Theater [color="/set19/1"]

```

```

Theater -> a_Arbeiter__und_Studententheater [color="/set19/1"]
Theater -> a_Besserungsstück [color="/set19/1"]
Theater -> a_Fabula_praetexta [color="/set19/1"]
Theater -> a_Fabula_togata [color="/set19/1"]
Theater -> a_Lehrstück__Theater_ [color="/set19/1"]
Theater -> a_Moralitäten [color="/set19/1"]
Theater -> a_Musical_Comedy [color="/set19/1"]
Theater -> a_Musical_Play [color="/set19/1"]
Theater -> a_Neidhartspiel [color="/set19/1"]
Theater -> a_Palliata [color="/set19/1"]
Theater -> a_Thingspiel [color="/set19/1"]
Theater -> a_Absurdes_Theater [color="/set19/2"]
Theater -> a_Arbeiter__und_Studententheater [color="/set19/2"]
Theater -> a_Besserungsstück [color="/set19/2"]
Theater -> a_Fabula_praetexta [color="/set19/2"]
Theater -> a_Fabula_togata [color="/set19/2"]
Theater -> a_Lehrstück__Theater_ [color="/set19/2"]
Theater -> a_Moralitäten [color="/set19/2"]
Theater -> a_Musical_Comedy [color="/set19/2"]
Theater -> a_Musical_Play [color="/set19/2"]
Theater -> a_Neidhartspiel [color="/set19/2"]
Theater -> a_Palliata [color="/set19/2"]
Theater -> a_Thingspiel [color="/set19/2"]

legende [label=<
<table cellpadding="4">
<tr><td>Legende</td></tr>
<tr><td><font color="/set19/1">UB_generisch_Form_Stoff</font></td>↓
></tr>
<tr><td><font color="/set19/2">UB_generisch_Zeit_historisch</font>↓
></td></tr>
</table>
>, shape=plaintext]

}

```

C.2 Weitere Beispiele

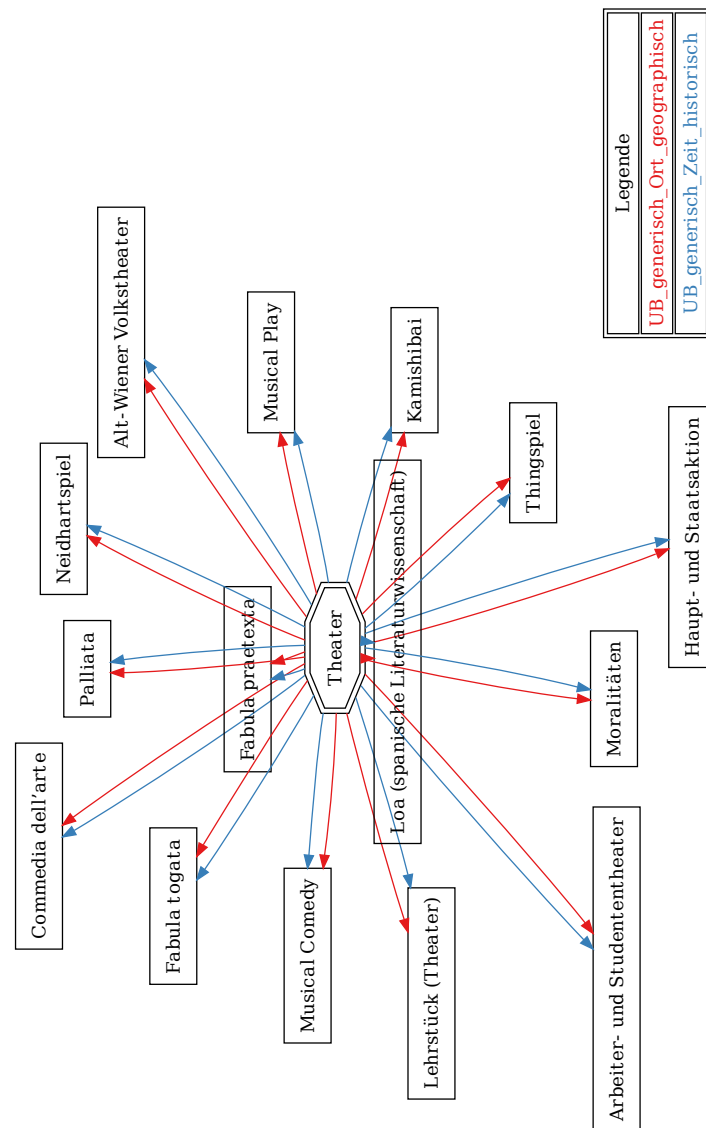


Abbildung 9: Die Unterbegriffe des Begriffs Theater, deren Unterordnung auf einem geographischen *und* einem zeitlichen Bezug gründet, in den neuen Strukturen

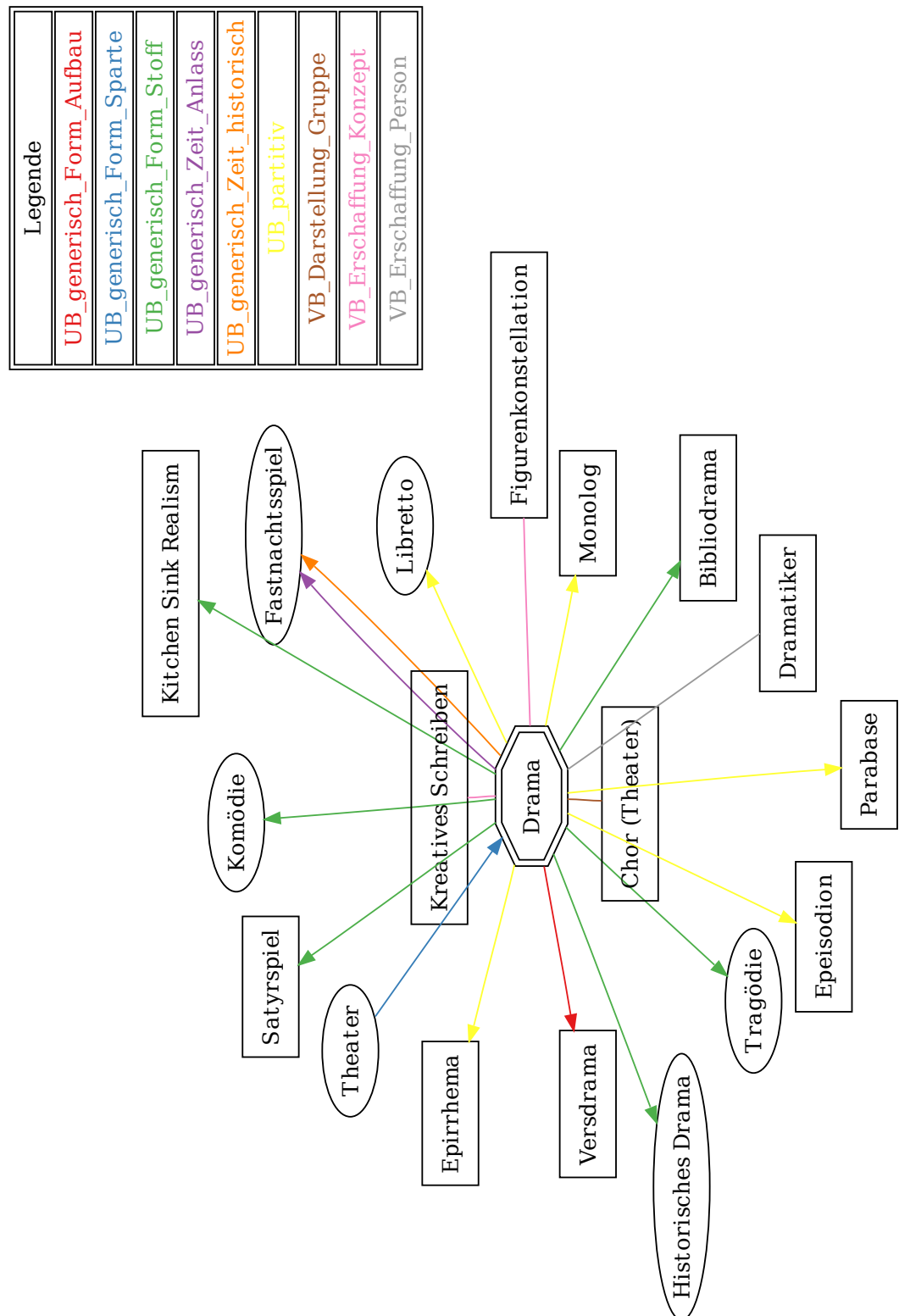


Abbildung 10: Das semantische Umfeld des Begriffs Drama in den neuen Strukturen (nur neun Relationstypen)

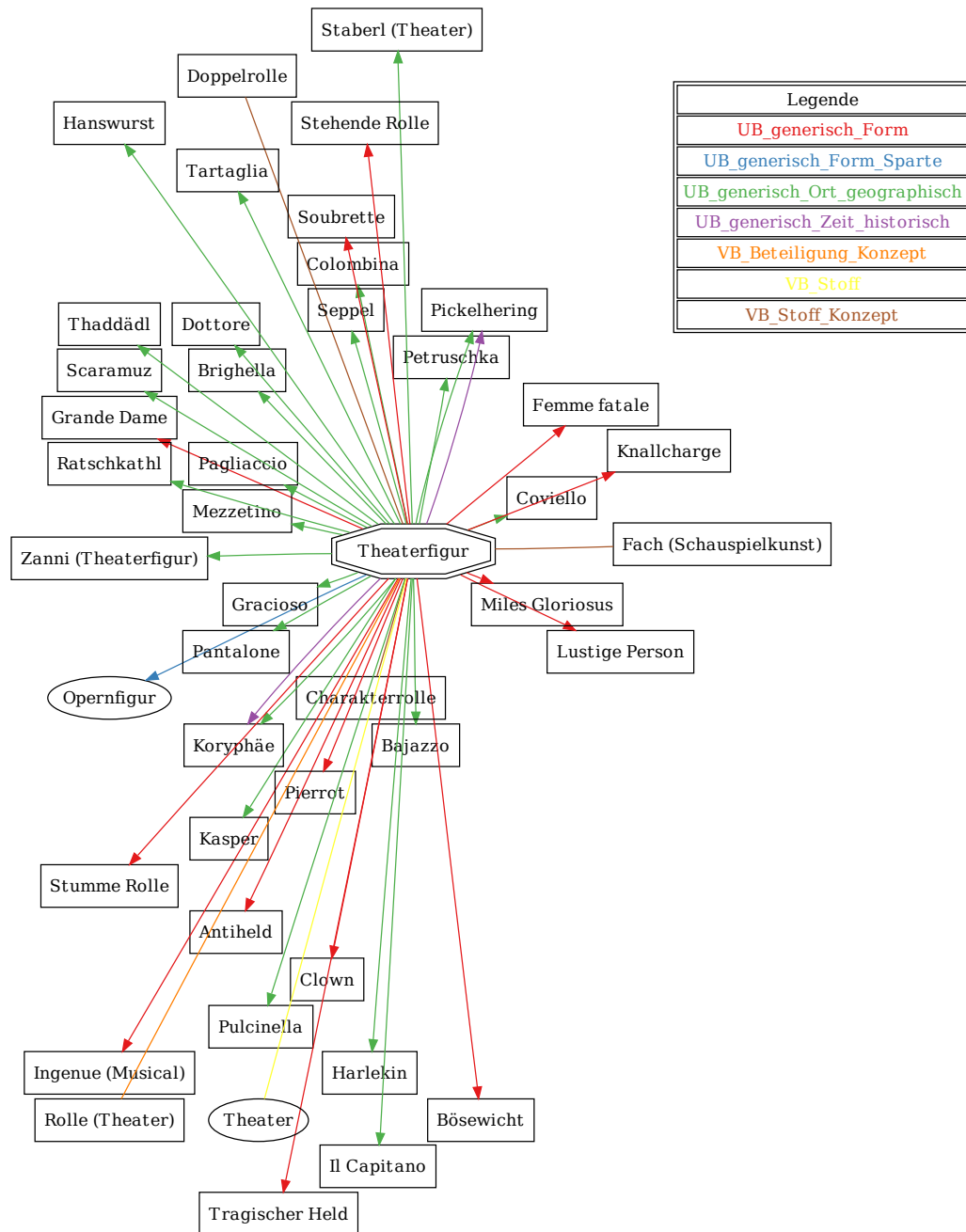


Abbildung 11: Das semantische Umfeld des Begriffs Theaterfigur in den neuen Strukturen

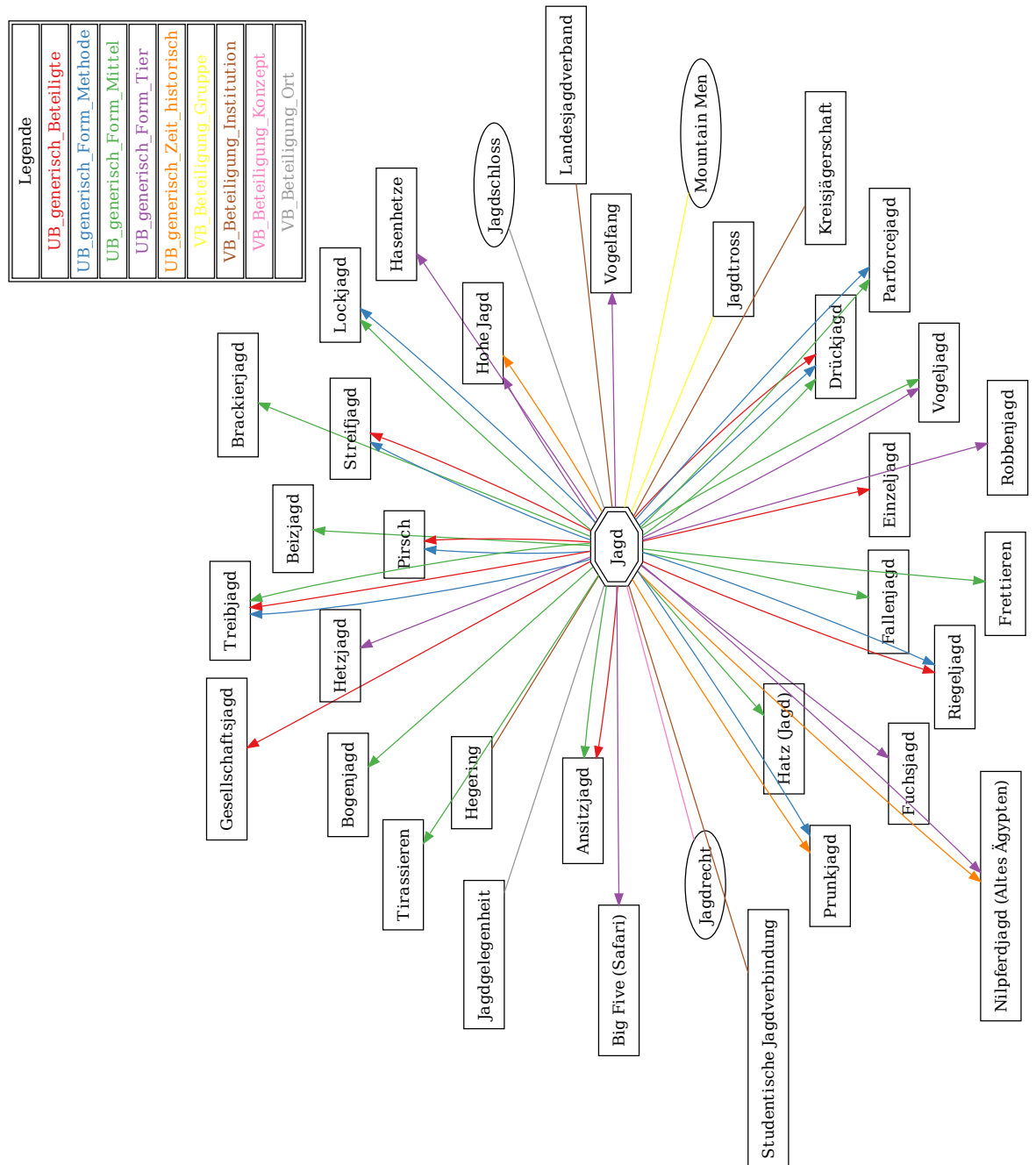


Abbildung 12: Das semantische Umfeld des Begriffs Jagd in den neuen Strukturen (nur neun Relationstypen)

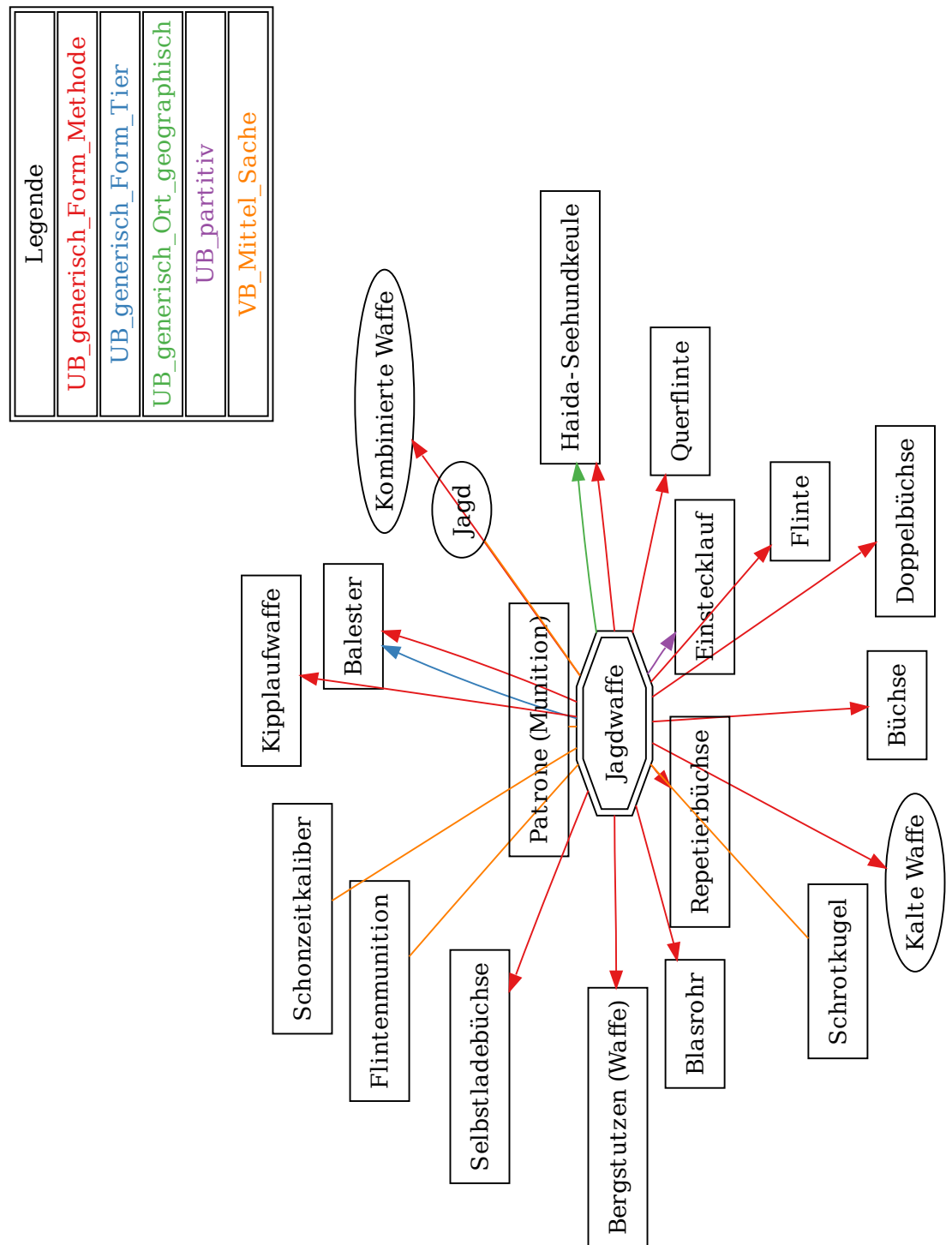


Abbildung 13: Das semantische Umfeld des Begriffs Jagdwaffe in den neuen Strukturen

D Skripte für „bash“

Es folgen die in den Abschnitten 5.3.3 beziehungsweise 5.3.5 beschriebenen Skripte in der Sprache der Unix-Shell „bash“. Sie sind auch auf der dieser Arbeit beigelegten CD-ROM im Verzeichnis `skripte/` enthalten.

Skript umarbeitung-dumps

```
#!/bin/bash

datum=20100603

if test $1; then

    if test "$1" = "categorylinks"
    then
        # page_id Titel_der_Kategorie
        zgrep '^INSERT INTO 'categorylinks' VALUES ' dumps/dewiki-↓
$datum-categorylinks.sql.gz | sed -e 's/^INSERT INTO '↓
categorylinks' VALUES //' -e 's/),(/)\
(/g' | sed "s/^(\([0-9]*\),'\([^']*'\).*$/\1 \2/" | gzip -c -- > ↓
categorylinks.gz

        elif test "$1" = "page"
        then
            # page_id namespace Titel Redirect
            # namespace: 0=Artikel, 14=Kategorie
            zgrep '^INSERT INTO 'page' VALUES ' dumps/dewiki-$datum-page.sql↓
.gz | sed -e 's/^INSERT INTO 'page' VALUES //' -e 's/),(/)\
(/g' | sed "s/^(\([0-9]*\),\([0-9]*\),'\([^']*'\)↓
', '[^']*',[0-9]*,\([0-9]*\).*$/\1 \2 \3 \4/" | egrep '^[0-9]*(\ (0)↓
| (14) )' | gzip -c -- > page.gz
# ineffizient

        fi

    else
        echo Benutzung: $0 \[page\|categorylinks\] 1>&2
    fi
fi
```

Skript browse-down

```
#!/bin/bash

if test $1; then
    directory=/Users/alex/Documents/FH/Bachelorarbeit/wikipedia
```

```

# $2 wird beim Selbstaufruf auf "n" gesetzt und darf
# ansonsten nicht besetzt sein.
if ! test $2; then
    rm -f /tmp/wikipedia-browse-down
    echo $1
fi

echo $1 >> /tmp/wikipedia-browse-down

prefix="a:"

alle_page_ids='zgrep " $1\$" $directory/categorylinks.gz | sed '↓
s/ .*//','
for page_id in $alle_page_ids; do
    page_line='zgrep "^$page_id " $directory/page.gz'
    page_name='echo $page_line | sed -e 's/^[0-9]* [0-9]* //' -e '↓
s/ .*//','
    page_namespace='echo $page_line | sed -e 's/^[0-9]* //' -e 's/↓
.*//','
    if test $page_namespace = 0; then
        echo $1 NT $prefix$page_name
    elif test $page_namespace = 14; then
        echo $1 NT $page_name
        if ! (grep "^$page_name" /tmp/wikipedia-browse-down > /dev/↓
null ); then
            $0 $page_name n
        fi
    fi
done

else
    echo Benutzung: $0 Name_der_Kategorie 1>&2
fi

```

Skript browse-up

```

#!/bin/bash

if test $2; then
    directory=/Users/alex/Documents/FH/Bachelorarbeit/wikipedia

    # $3 wird beim Selbstaufruf auf "n" gesetzt und darf
    # ansonsten nicht besetzt sein.
    if ! test $3; then
        rm -f /tmp/wikipedia-find-page
    fi

```

```

echo $1 $2 >> /tmp/wikipedia-find-page

page_name=$1
page_namespace=$2
page_line='zgrep " $page_namespace $page_name " $directory/page.↓
gz'

if ! test "$page_line"; then
    if test $page_namespace = 0; then
        # Wenn 0 nicht gefunden, dann mit 14 versuchen.
        page_line='zgrep " 14 $page_name " $directory/page.gz'
        page_namespace=14
    fi
fi

prefix=
if test $page_namespace = 0; then
    prefix="a:"
fi

if test "$page_line"; then

    page_id='echo $page_line | sed 's/ .*//','
    is_redirect='echo $page_line | sed 's/^.* //'','
    if test $is_redirect = 1; then
        redirect_line='zgrep "^$page_id " $directory/redirect.gz'
        redirect_name='echo $redirect_line | sed 's/^.* //'','
        redirect_namespace='echo $redirect_line | sed -e 's/^[0-9]* ↓
//' -e 's/ .*//','
        page_id='zgrep " $redirect_namespace $redirect_name " ↓
$directory/page.gz | sed 's/ .*//','
        echo $redirect_name $redirect_namespace >> /tmp/wikipedia-↓
find-page
        redirect_prefix=
        if test $redirect_namespace = 0; then
            redirect_prefix="a:"
        fi
        echo $prefix$page_name USE $redirect_prefix$redirect_name
        page_name=$redirect_name
        prefix=$redirect_prefix
    fi

    all_cat_names='zgrep "^$page_id " $directory/categorylinks.gz ↓
| sed 's/^[0-9]* //'','
    if ! test "$all_cat_names";
    then
        echo $page_name TT
    fi
fi

```

```

fi
for cat_name in $all_cat_names; do
    #cat_page_id='zgrep " 14 $cat_name " $directory/page.gz | ↓
sed 's/ .*/'/ '
    echo $prefix$page_name BT $cat_name
    if ! ( grep "^$cat_name 14" /tmp/wikipedia-find-page > /dev/↓
null ); then
        $0 $cat_name 14 n
    fi
done

else
    echo $prefix$page_name ??
fi

else
    echo Benutzung: $0 Name Namespace 1>&2
fi

```

Skript tree-top-down-plain

```

#!/bin/bash

prefix="a:"

if test $1; then
    filename="$1"
    if test "$2"; then
        # Topsterme wurden angegeben
        topterms="$2"
    else
        # Suche Topsterme
        # = Term x, fuer den nicht gilt:
        # x BT y
        # y NT x
        all_terms='grep -v ' USE ' $filename | egrep -v '^[^ ]+ ..$' |
sed 's/ [A-Z][A-Z] / /' | tr ' ' '\n' | sort | uniq'
        for at in $all_terms; do
            if ! ( grep "^$at BT" $filename > /dev/null ); then
                if ! ( grep " NT $at\$" $filename > /dev/null ); then
                    topterms="$topterms $at"
                fi
            fi
        done
    fi

    # $3 wird beim Selbstaufruf auf "n" gesetzt und darf

```

```

# ansonsten nicht besetzt sein.
if ! test $3; then
    rm -f /tmp/tree-top-down-names
fi

for term in $topterms; do

    if ! test $3; then
        tt_not_found=0
        if ( egrep '(\<' $term '\>)|(\<' $term '\>$)' $filename > /dev/↓
null ); then
            echo $term TT
        fi
    fi

    echo $term >> /tmp/tree-top-down-names

    alle_nt_via_bt='grep ' BT ' $term '$' $filename | sed 's/ BT .*$↓
//','
    alle_nt_via_nt='grep ^$term' NT' $filename | sed 's/^.* NT ↓
//','
    alle_nt='echo $alle_nt_via_bt $alle_nt_via_nt | tr ' ' '\n' | ↓
sort | uniq'
    if test "$alle_nt"; then
        for nt in $alle_nt; do
            if ( grep "^$nt\$" /tmp/tree-top-down-names > /dev/null );↓
then
                echo "$term UB $nt (s.o.)"
            else
                echo $term UB $nt
                if ! ( echo $nt | grep "^$prefix" > /dev/null ); then
                    $0 $filename $nt n
                fi
            fi
        done
    fi

done

if ! test $3; then
    non_existing='grep ' ?? '$' $filename | sed 's/ ??$/','
    if test "$non_existing"; then
        for non in $non_existing; do
            echo "$non existiert nicht."
        done
    fi
fi
fi

```

```

else
    echo Benutzung: $0 Dateiname \[Topterm\] 1>&2
fi

```

Skript tree-top-down-html

```

#!/bin/bash

prefix="a:"

if test $1; then
    filename="$1"
    if test "$2"; then
        # Topterme wurden angegeben
        topterm="$2"
    else
        # Suche Topterme
        # = Term x, fuer den nicht gilt:
        # x BT y
        # y NT x
        all_terms='grep -v 'USE' $filename | egrep -v '^[^ ]+ ..$' |
sed 's/[A-Z][A-Z] / /' | tr ' ' '\n' | sort | uniq'
        for at in $all_terms; do
            if ! ( grep "^$at BT" $filename > /dev/null ); then
                if ! ( grep " NT $at\$" $filename > /dev/null ); then
                    topterm="$topterm $at"
                fi
            fi
        done
    fi

    # $3 wird beim Selbstaufruf auf "n" gesetzt und darf
    # ansonsten nicht besetzt sein.
    if ! test $3; then
        rm -f /tmp/tree-top-down-names
        echo '<meta http-equiv="content-type" content="text/html; ↓
charset=UTF-8">'
    fi

    for term in $topterm; do

        if ! test $3; then
            tt_not_found=0
            if ! ( egrep '(\<'$term'\>)|(\<'$term'\>$)' $filename > /↓
dev/null ); then
                tt_not_found=1
                echo "<p>Topterm $term nicht gefunden.</p>"
            fi
        fi
    done

```



```

        else
            echo "<ul>"
            echo "<li>$term</li>"
        fi
    fi

    echo $term >> /tmp/tree-top-down-names

    alle_nt_via_bt='grep ' BT '$term'$' $filename | sed 's/ BT .*$↓
//','
    alle_nt_via_nt='grep ^$term' NT' $filename | sed 's/^.* NT ↓
//','
    alle_nt='echo $alle_nt_via_bt $alle_nt_via_nt | tr ' ' '\n' | ↓
sort | uniq'
    if test "$alle_nt"; then
        echo "<ul>"
        for nt in $alle_nt; do
            if ( grep "^$nt\$" /tmp/tree-top-down-names > /dev/null );↓
then
                echo "<li><a href=\"#$nt\">$nt</a></li>"
            else
                if ( echo $nt | grep "^$prefix" > /dev/null ); then
                    nt_ohne_prefix='echo $nt | sed "s/^$prefix/"'
                    echo "<li><a name=\"$nt\" href=\"http://de.wikipedia.↓
org/wiki/$nt_ohne_prefix\">$nt</a></li>"
                else
                    echo "<li><a name=\"$nt\" href=\"http://de.wikipedia.↓
org/wiki/Kategorie:$nt\">$nt</a></li>"
                    $0 $filename $nt n
                fi
            fi
        done
        echo "</ul>"
    fi

    if ! test $3; then
        if test $tt_not_found = 0; then
            echo "</ul>"
        fi
    fi

done

if ! test $3; then
    non_existing='grep ' ??'$' $filename | sed 's/ ??$/','
    if test "$non_existing"; then
        for non in $non_existing; do
            echo "<p>$non existiert nicht.</p>"
        done
    fi
fi

```

```

        done
    fi
fi

else
    echo Benutzung: $0 Dateiname \[Topterm\] 1>&2
fi

```

Skript visualisierung-or

```

#!/bin/bash

rm -f /tmp/visualisierung-or-*
sed_escape='s/[()\\:\-]/_/g'

if test $2; then
    filename=$1; shift
    term=$1; shift
    relationen="$*"
    if test $# \> 9
    then
        echo Es koennen nur maximal 9 Relationstypen verarbeitet ↓
        werden. 1>&2
    fi

    grep "^$term " "$filename" > /tmp/visualisierung-or-alle-ab
    grep " $term\$" "$filename" > /tmp/visualisierung-or-alle-zu

    if ! test "$relationen"
    then
        relationen='cat /tmp/visualisierung-or-alle-* | awk '{ print ↓
$2 }' | sort | uniq | grep -v '_Instanz$' | grep -v '^--$' | grep ↓
-v '^TT$' | head -n 9'
    fi

    for r in `echo $relationen`
    do
        grep "^$term $r " /tmp/visualisierung-or-alle-ab | awk '{ ↓
print $3 }' > /tmp/visualisierung-or-relation-ab-$r
        grep " $r $term$" /tmp/visualisierung-or-alle-zu | awk '{ ↓
print $1 }' > /tmp/visualisierung-or-relation-zu-$r
    done

    echo "# Vokabular aus Datei $filename"
    echo "# Begriff: $term"
    echo "# Relationstypen: "`echo $relationen`
    echo "# OR-Verknuepfung"

```

```

echo
echo "digraph $term {"
echo
term_anzeige='echo $term | sed 's/_/ /g','
echo $term "[label=\"${term_anzeige}\", shape=doubleoctagon]"
echo

for t in `cat /tmp/visualisierung-or-relation-* | sort | uniq`
do
    t_anzeige='echo $t | sed -e 's/^a:/' -e 's/_/ /g','
    if ( echo $t | grep "^a:" > /dev/null )
    then
        echo `echo $t | sed $sed_escape` "[label=\"${t_anzeige}\" ↓
shape=box]"
    else
        echo $t "[label=\"${t_anzeige}\" shape=ellipse]"
    fi
done

rnum=0
for r in `echo $relationen`
do
    rnum=`expr $rnum + 1`
    direction=
    if ( echo $r | grep "VB_" > /dev/null )
    then
        direction=" dir=none"
    fi
    for t in `cat /tmp/visualisierung-or-relation-ab-$r`
    do
        echo $term "->" `echo $t | sed $sed_escape` "[color=\"/set19↓
/$rnum\"$direction]"
    done
    for t in `cat /tmp/visualisierung-or-relation-zu-$r`
    do
        echo `echo $t | sed $sed_escape` "->" $term "[color=\"/set19↓
/$rnum\"$direction]"
    done
done

echo
echo `legende [label=<`
echo `<table cellpadding="4">`
echo `<tr><td>Legende</td></tr>`

rnum=0
for r in `echo $relationen`
do

```

```

        rnum='expr $rnum + 1'
        echo "<tr><td><font color=\" /set19/$rnum\">$r</font></td></tr>↓"
    "
done

echo '</table>'
echo '>, shape=plaintext]'
echo
echo '}'

else
    echo Benutzung: $0 Dateiname Begriff \[Relationstyp ...\] 1>&2
    echo Im Begriff und in den Relationstypen sind Leerzeichen 1>&2
    echo durch Unterstriche \(_\) zu ersetzen. 1>&2
    echo Wird kein Relationstyp angegeben, so werden die ersten 9 ↓
    aller 1>&2
    echo vorkommenden Relationstypen mit Ausnahme aller ↓
    Instanzrelationen 1>&2
    echo dargestellt. 1>&2
fi

```

Skript visualisierung-and

```

#!/bin/bash

rm -f /tmp/visualisierung-and-*
sed_escape='s/[()\\:\-]/_/g'

if test $2; then
    filename=$1; shift
    term=$1; shift
    relationen="$*"
    if test $# \> 9
    then
        echo Es koennen nur maximal 9 Relationstypen verarbeitet ↓
        werden. 1>&2
    fi

    grep "^$term " "$filename" > /tmp/visualisierung-and-alle-ab
    grep " $term\$" "$filename" > /tmp/visualisierung-and-alle-zu

    if ! test "$relationen"
    then
        relationen='cat /tmp/visualisierung-and-alle-* | awk '{ print ↓
        $2 }' | sort | uniq | grep -v '_Instanz$' | grep -v '^--$' | grep ↓
        -v '^TT$' | head -n 9'
    fi
fi

```

```

fi

for r in `echo $relationen`
do
    grep "^$term $r " /tmp/visualisierung-and-alle-ab >> /tmp/↓
visualisierung-and-relationen-ab
    grep " $r $term$" /tmp/visualisierung-and-alle-zu >> /tmp/↓
visualisierung-and-relationen-zu
done

terme_ab=`cat /tmp/visualisierung-and-relationen-ab | awk '{ ↓
print $3 }' | sort | uniq -c | grep "^ *$# " | awk '{ print $2 }'`
terme_zu=`cat /tmp/visualisierung-and-relationen-zu | awk '{ ↓
print $1 }' | sort | uniq -c | grep "^ *$# " | awk '{ print $2 }'`

echo "# Vokabular aus Datei $filename"
echo "# Begriff: $term"
echo "# Relationstypen: "`echo $relationen`
echo "# AND-Verknuepfung"
echo
echo "digraph $term {"
echo
term_anzeige=`echo $term | sed 's/_/ /g'`
echo $term "[label=\"$term_anzeige\", shape=doubleoctagon]"
echo

for t in `echo $terme_ab $terme_zu`
do
    t_anzeige=`echo $t | sed -e 's/^a:/' -e 's/_/ /g'`
    if ( echo $t | grep "^a:" > /dev/null )
    then
        echo `echo $t | sed $sed_escape` "[label=\"$t_anzeige\" ↓
shape=box]"
    else
        echo $t "[label=\"$t_anzeige\" shape=ellipse]"
    fi
done

rnum=0
for r in `echo $relationen`
do
    rnum=`expr $rnum + 1`
    direction=
    if ( echo $r | grep "VB_" > /dev/null )
    then
        direction=" dir=none"
    fi
    for t in `echo $terme_ab`

```

```

do
    echo $term "->" 'echo $t | sed $sed_escape' "[color=\"/set19↓
/$rnum\"$direction]"
done
for t in 'echo $terme_zu'
do
    echo 'echo $t | sed $sed_escape' "->" $term "[color=\"/set19↓
/$rnum\"$direction]"
done
done

echo
echo 'legende [label=<'
echo '<table cellpadding="4">'
echo '<tr><td>Legende</td></tr>'

rnum=0
for r in 'echo $relationen'
do
    rnum='expr $rnum + 1'
    echo "<tr><td><font color=\"/set19/$rnum\">$r</font></td></tr>↓
"
done

echo '</table>'
echo '>, shape=plaintext]'
echo
echo '}'

else
    echo Benutzung: $0 Dateiname Begriff \[Relationstyp ...\] 1>&2
    echo Im Begriff und in den Relationstypen sind Leerzeichen 1>&2
    echo durch Unterstriche \(_\) zu ersetzen. 1>&2
    echo Wird kein Relationstyp angegeben, so werden die ersten 9 ↓
    aller 1>&2
    echo vorkommenden Relationstypen mit Ausnahme aller ↓
    Instanzrelationen 1>&2
    echo dargestellt. 1>&2
fi

```

E Verzeichnisstruktur der beigelegten CD-ROM

Verzeichnisname

<code>theater/</code>	enthält die Datei <code>Theater-bearbeitet.txt</code> mit der neuen Vokabularstruktur sowie die anderen im Text erwähnten Dateien zum Gegenstandsbereich Theater.
<code>jagd/</code>	enthält die Datei <code>Jagd-bearbeitet.txt</code> mit der neuen Vokabularstruktur sowie die anderen im Text erwähnten Dateien zum Gegenstandsbereich Jagd.
<code>skripte/</code>	enthält sämtliche „bash“-Skripte.
<code>visualisierung/</code>	enthält sämtliche Visualisierungsbeispiele als PDF-Dateien sowie deren Quelltexte im Graphviz-Format.

Eidesstattliche Erklärung

Hiermit versichere ich, die Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt zu haben.

Köln, den

Alexander Meyer